



# 20



**НАЦИОНАЛЬНАЯ  
КОНФЕРЕНЦИЯ  
ПО ИСКУССТВЕННОМУ  
ИНТЕЛЛЕКТУ  
С МЕЖДУНАРОДНЫМ  
УЧАСТИЕМ**



## **КИИ-2022**



**ТРУДЫ КОНФЕРЕНЦИИ**

**Том 1**

21-23 декабря 2022 г.  
Москва

# 20

Министерство науки  
и высшего образования РФ

Российская ассоциация  
искусственного интеллекта

Федеральный  
исследовательский центр  
«Информатика и  
управление» РАН

Национальный  
исследовательский  
университет «МЭИ»

## НАЦИОНАЛЬНАЯ КОНФЕРЕНЦИЯ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ С МЕЖДУНАРОДНЫМ УЧАСТИЕМ

### КИИ-2022

ТРУДЫ КОНФЕРЕНЦИИ

Том 1

21–23 декабря 2022 г.  
Москва, НИУ «МЭИ»

*К 90-летию со дня рождения  
Дмитрия Александровича Поспелова*

Москва  
Издательство МЭИ  
2022

УДК 004.8+004.89+004.82+004.032.26(045)

ББК 32.813

Д 22

### Организаторы конференции:

*Российская ассоциация искусственного интеллекта  
Федеральный исследовательский центр «Информатика и управление» РАН  
Национальный исследовательский университет «МЭИ»*

Рецензенты: академик РАН, ИПУ РАН С.Н. Васильев,  
д.т.н., доц., ФИЦ ИУ РАН В.Н. Захаров

**Д 22 Двадцатая Национальная конференция по искусственному интеллекту с международным участием, КИИ-2022** (Москва, 21–23 декабря 2022г.). Труды конференции. В 2 т. Т. 1. – М.: Издательство МЭИ, 2022. – 376 с.

**ISBN 978-5-7046-2735-7**

**ISBN 978-5-7046-2736-4 (Т. 1)**

Юбилейная двадцатая конференция КИИ-2022 проходит в год 90-летия со дня рождения Дмитрия Александровича Поспелова – выдающегося отечественного ученого в области искусственного интеллекта, одного из главных организаторов этого направления исследований и разработок в России.

В первом томе трудов публикуются пленарные доклады, а также доклады участников конференции, представленные на следующих секциях:

Секция 1 «Инженерия знаний, когнитивные исследования и интеллектуальный анализ текстов»,

Секция 2 «Интеллектуальный анализ данных и объяснимый искусственный интеллект».

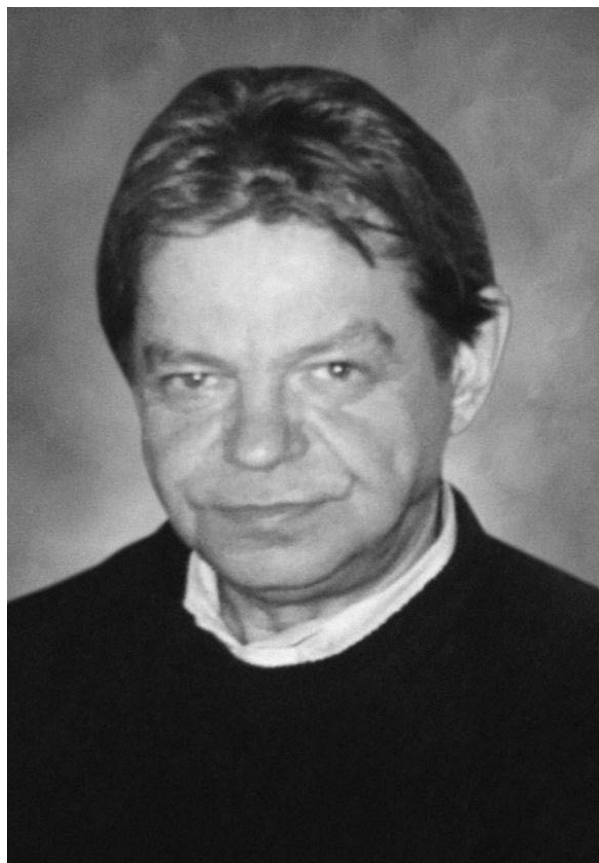
**УДК 004.8+004.89+004.82+004.032.26(045)  
ББК 32.813**

*Материалы печатаются в авторской редакции методом  
прямого воспроизведения с авторских оригиналов*

ISBN 978-5-7046-2735-7

ISBN 978-5-7046-2736-4 (Т. 1)

© Национальный исследовательский университет «МЭИ», 2022



*Дмитрий Александрович Поспелов  
(1932–2019 гг.)*

## СОПРЕДСЕДАТЕЛИ КОНФЕРЕНЦИИ

Соколов И.А., акад. РАН, ФИЦ ИУ РАН, Москва

Роголёв Н.Д., д.т.н., проф., НИУ «МЭИ», Москва

## ПРОГРАММНЫЙ КОМИТЕТ КОНФЕРЕНЦИИ

### *Председатель Программного комитета*

Кобринский Б.А., д.м.н., проф., ФИЦ ИУ РАН, Москва

### *Заместители председателя Программного комитета*

Еремеев А.П., д.т.н., проф., НИУ «МЭИ», Москва

Забейало М.И., д.ф.-м.н., проф., ФИЦ ИУ РАН, Москва

### *Ответственный секретарь Программного комитета*

Виноградов Д.В., д.ф.-м.н., ФИЦ ИУ РАН, Москва

### *Члены программного комитета*

Аверкин А.Н., к.ф.-м.н., доц., ФИЦ ИУ РАН, Москва

Алиев Р.А., д.т.н., проф., АЗГНА, Азербайджан, Баку

Афанасьева Т.В., д.т.н., РЭУ им. Г.В. Плеханова, Москва

Базенков Н.И., к.т.н., с.н.с. ИПУ РАН, Москва

Борисов В.В., д.т.н., проф., филиал НИУ «МЭИ», Смоленск

Васильев С.Н., акад. РАН, ИПУ РАН, г. Москва

Гаврилова Т.А., д.т.н., проф., СПбГУ, Санкт-Петербург

Городецкий В.И., д.т.н., проф., АО «Эврика», Санкт-Петербург

Грибова В.В., член-корр. РАН, ИАПУ ДВО РАН, Владивосток

Драгунов В.К., д.т.н., проф., НИУ «МЭИ», Москва

Добрынин Д.А., к.т.н., с.н.с. ФИЦ ИУ РАН, Москва

Жилиякова Л.Ю., д.ф.-м.н., в.н.с. ИПУ РАН, Москва

Загорюлько Ю.А., к.т.н., ИСИ СО РАН, Новосибирск

Захаров В.Н., д.т.н., доц., ФИЦ ИУ РАН, Москва

Зацаринный А.А., д.т.н., доц., ФИЦ ИУ РАН, Москва

Ковалев С.М., д.т.н., проф., РГУПС, Ростов-на-Дону

Колесников А.В., д.т.н., проф., БФУ, Калининград

Колоденкова А.Е., д.т.н., доц., СамГТУ, Самара

Котенко И.В., д.т.н., проф., СПИИРАН, СПб

Кузнецов О.П., д.т.н., проф., ИПУ РАН, Москва

Кузнецов С.О., д.ф.-м.н., проф., НИУ ВШЭ, Москва

Кулинич А.А., к.т.н., доц., ИПУ РАН, Москва

Курейчик В.В., д.т.н., проф., ЮФУ, Ростов-на-Дону

Лебедев О.Б., д.т.н., доц., ИКТИБ ЮФУ, Таганрог

Лукашевич Н.В., д.т.н., проф. НИВЦ МГУ, Москва

Мисник А.Е., к.т.н., доц., БРУ, Республика Беларусь, Могилев

Михеенкова М.А., д.т.н., проф., ФИЦ ИУ РАН, Москва

Пальчунов Д.Е., д.ф.-м.н., доц., ИМ СО РАН, Новосибирск  
Паллох Б.В., д.т.н., проф., ТвГТУ, Тверь  
Панов А.И., к.ф.-м.н., доц., ФИЦ ИУ РАН, Москва  
Редько В.Г., д.ф.-м.н., проф., НИИСИ РАН, Москва  
Ройзензон Г.В., к.т.н., доц., ФИЦ ИУ РАН, Москва  
Рыбина Г.В., д.т.н., проф., НИЯУ МИФИ, Москва  
Семенов Н.А., д.т.н., проф. ТвГТУ, Москва  
Смирнов И.В., к.ф.-м.н., доц., ФИЦ ИУ РАН, Москва  
Стефанюк В.Л., д.т.н., проф., ИППИ РАН, Москва  
Сулейманов Д.Ш., акад. АН РТ, ИПС АН РТ, Казань  
Тельнов Ю.Ф., д.э.н., проф., РЭУ, Москва  
Федунов Б.Е., д.т.н., проф., ГосНИИ АС, Москва  
Финн В.К., д.т.н., проф., ФИЦ ИУ РАН, г. Москва  
Фоминых И.Б., д.т.н., проф., НИУ «МЭИ», г. Москва  
Харламов А.А., д.т.н., проф., ИВНД и НФ РАН, Москва  
Хачумов В.М., д.т.н., проф., ФИЦ ИУ РАН, Москва  
Хорошевский В.Ф., д.т.н., проф., ФИЦ ИУ РАН, Москва  
Юдин Д.А. к.т.н., с.н.с. МФТИ, Москва  
Юсупбеков Н.Р., акад. АН РУз, ТашГТУ, Узбекистан, Ташкент  
Яковлев К.С., к.ф.-м.н., ФИЦ ИУ РАН, Москва  
Ярушкина Н.Г., д.т.н., проф., УлГТУ, Ульяновск

## **ОРГАНИЗАЦИОННЫЙ КОМИТЕТ КОНФЕРЕНЦИИ**

### *Председатель Организационного комитета*

Борисов В.В., д.т.н., проф., филиал НИУ «МЭИ», г. Смоленск

### *Заместитель председателя Организационного комитета*

Бобряков А.В., д.т.н., проф., НИУ «МЭИ», г. Москва

### *Члены организационного комитета*

Варшавский П.Р., к.т.н., доц., НИУ «МЭИ», г. Москва

Вишняков С.В., к.т.н., доц. НИУ «МЭИ», г. Москва

Гладков Л.А., к.т.н., доц., ИКТИБ ЮФУ, г. Таганрог

Луферов В.С., к.т.н., филиал НИУ «МЭИ», г. Смоленск

Синявский Ю.В., к.т.н., доц., филиал НИУ «МЭИ», г. Смоленск

Солопов Р.В., к.т.н., доц., филиал НИУ «МЭИ», г. Смоленск

## **ИНФОРМАЦИОННАЯ ПОДДЕРЖКА КОНФЕРЕНЦИИ**

ООО «Лаборатория информационных технологий» (Россия, Смоленск)

## ПРЕДИСЛОВИЕ

Юбилейная XX Национальная конференция по искусственному интеллекту с международным участием (КИИ-2022) своими истоками восходит к первой КИИ в 1988 году и продолжает традицию советских (российских) конференций, у основания которых стоял Дмитрий Александрович Поспелов – выдающийся отечественный ученый в области искусственного интеллекта, 90-летие со дня рождения которого мы отмечаем в 2022 году.

Организаторами конференции являются Российская ассоциация искусственного интеллекта, Федеральный исследовательский центр «Информатика и управление» РАН, Национальный исследовательский университет «МЭИ».

Тематика конференции охватывает следующие основные направления искусственного интеллекта: инженерия знаний; когнитивные исследования; интеллектуальный анализ текстов; интеллектуальный анализ данных; объяснимый искусственный интеллект; нейросетевые и нечеткие методы, мягкие измерения и вычисления; интеллектуальные системы поддержки принятия решений и управления; прикладные интеллектуальные системы и инструментальные средства построения интеллектуальных систем.

На конференцию представлены пленарные доклады видных ученых и специалистов в области искусственного интеллекта. Из поданных на конференцию 89 секционных докладов Программным комитетом после рецензирования отобраны 59 докладов.

В конференции принимают участие более 150 ученых, исследователей и специалистов из академических институтов, университетов, отраслевых институтов и организаций Москвы, Санкт-Петербурга, Белгорода, Владивостока, Иркутска, Казани, Калининграда, Новосибирска, Ростова-на-Дону, Самары, Смоленска, Таганрога, Твери, Ульяновска, Уфы, Могилёва (Республика Беларусь).

Мы посвящаем это издание памяти наших дорогих коллег – Дмитрия Александровича Поспелова и многолетнего члена Научного совета Российской ассоциации искусственного интеллекта Делира Гасемовича Лахути, ушедшего из жизни в этом году.

Делир Гасемович Лахути (1934–2022 гг.) – доктор технических наук, профессор, выдающийся специалист в области информатики, философии, филологии. Круг его научных интересов включал в себя автоматический анализ текстов на естественном языке и информационно-поисковые системы полнотекстовых баз данных. Он – автор уникальных работ о проблемах перевода философских текстов. Его книга «Вдумываясь в текст» – оригинальное исследование точной филологии. Человек выдающейся эрудиции, Д.Г. Лахути был одним из лучших переводчиков научных текстов, чьи переводы книг Л. Витгенштейна, У.Р. Эшби, Ч.С. Пирса, К.Р. Поппера считаются классическими.

Выражаем искреннюю благодарность Д.В. Виноградову, В.С. Луферову, Н.В. Ярошевич, Ю.В. Синявскому, Р.В. Солопову, Е.В. Вельмакиной за активное участие в подготовке трудов к печати, в организации и проведении конференции.

*Программный и Организационный комитеты КИИ-2022*



---

# Пленарные доклады

УДК 005.94

## ЗНАНИЯ И ДАННЫЕ В ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ: ДУЭЛЬ ИЛИ ДУЭТ\*

Т.А. Гаврилова (*gavrilova@gsom.spbu.ru*)

Высшая школа менеджмента СПбГУ, Санкт-Петербург

В докладе обсуждаются методологические аспекты двух основных направлений в области искусственного интеллекта – нейросетевого и семантического. Рассматриваются базовые парадигмы индуктивного и дедуктивного подхода, а также специфика данных и знаний как «топлива» интеллектуальных технологий.

**Ключевые слова:** структурирование знаний, онтологии, графы знаний, когнитивные стили.

### Введение

Расширяется повсеместное внедрение искусственного интеллекта (ИИ) в качестве неотъемлемой части продуктов, услуг и решений в современном обществе [Floridi et al., 2018; AI, 2019; Джонс, 2019]. Этот тренд ускоряет создание специализированных моделей ИИ, которые могут применяться для автоматизации производственных процессов, разработки нового программного обеспечения, обучения, различных прогнозов, машинного перевода и генерации текстов и изображений. ИИ прошел уже длинный путь от момента своего рождения в 1956 г. до настоящего времени [Stottler, 2002; AIT, 2022; Zhang & Lu, 2021; Захаров, Попов, Поспелов, & Хорошевский, 1987; Кобринский, 2022; Осипов, 2022; Финн, 2020]. При этом параллельно живут два мира ИИ, появившихся с середины 50-х.

---

\* Работа выполнена при частичной финансовой поддержке РФФИ (проект № 20-07-00854).

Вскоре после старта первых исследований в области искусственного интеллекта произошло разделение этой научной области на два направления: нейрокибернетика (индуктивный подход на основании данных) и «символьный искусственный интеллект» (дедуктивный подход на основании знаний). Эти направления развиваются практически независимо, существенно различаясь как в методологии, так и в технологиях. При этом попыток их сближения было достаточно много и они продолжают происходить (все эти, strong AI, explainable AI и AGI). В свете оглушительных успехов нейросетей возникает потребность осмысления основных различий и противоречий в этих двух ветвях ИИ (рис. 1).



*Рис. 1. Две ветви искусственного интеллекта*

Данный доклад представляет попытку такого осмысления на основании многолетнего опыта автора в области символического ИИ, обзора современной литературы и обширного рынка приложений. Ни философских, ни исторических основ ИИ автор намеренно не рассматривает в связи с необозримостью и глубиной темы, и ограничениями собственных познаний. Доклад – это прагматический взгляд с позиций information science.

## **1. Борьба двух парадигм**

Сначала о лидере. Основная идея нейро-подхода была сформулирована следующим образом: единственный объект, способный мыслить, — это человеческий мозг, поэтому любое «мыслящее» устройство должно каким-то образом воспроизводить его структуру.

Таким образом, нейрокибернетика ориентирована на программно-аппаратное моделирование структур, подобных структуре мозга [Горбань 2002]. Физиологами давно установлено, что основой человеческого мозга

является большое количество (до  $10^{21}$ ) связанных между собой и взаимодействующих нервных клеток – нейронов. Поэтому усилия нейрокибернетики были сосредоточены на создании элементов, аналогичных нейронам и их объединении в функционирующие системы. Эти системы принято называть искусственными *нейронными сетями*, или *нейросетями*.

Первые нейросети были созданы Розенблаттом и Мак-Каллоком в 1956–1965 гг. Это были попытки создать системы, моделирующие человеческий глаз и его взаимодействие с мозгом. Устройство, созданное ими тогда, получило название персептрона (perceptron). Персептрон умел различать буквы алфавита, но был слишком чувствителен к их написанию. Например, буквы “А”, “~~А~~” и “а” для этого устройства были тремя разными знаками. Постепенно в 70–80-х гг. количество работ в этом направлении искусственного интеллекта стало снижаться. Слишком неутешительны были первые результаты. Авторы объясняли неудачи малой памятью и низким быстродействием существующих в то время компьютеров. Ситуация изменилась в связи с технологическим прорывом и сверхвозможностями современных процессоров. Со временем нейросети стали «учиться» все лучше, и успехи этого направления сегодня трудно переоценить, что успешно демонстрируют, например, картины, созданные алгоритмом GAN (см. <https://www.artaigallery.com/>) или успехи в распознавании лиц и голоса. Нейросети успешно создают тексты, рисуют картины по текстовым описаниям и имитируют разговор [Горбань, 2000; Ни-коленко и др. 2018; Оселедец, 2022].

В основе этого лежат модели с миллионом параметров и огромные массивы выборок для обучения. Голая математика и никаких признаков «интеллекта» даже в вульгарном понимании этого слова. На вопрос “почему” эти алгоритмы также ответить не могут.

В основе этой парадигмы лежит статистическая обработка данных, очень больших данных в буквальном понимании этого слова.

Второе направление – символичный ИИ, или кибернетика «черного ящика». В основу подхода, названного кибернетикой «черного ящика», был положен принцип, противоположный нейрокибернетике: «Не имеет значения, как устроено “мыслящее” устройство. Главное, чтобы на заданные входные воздействия оно реагировало так же, как человеческий мозг». Для того, чтобы понять центральную парадигму этого подхода – работу со знаниями, как закономерностями и принципами функционирования, надо обратиться к истории.

Сторонники этого направления мотивировали свой подход тем, что человек не должен слепо следовать природе в своих научных и технологических поисках, тем более в аспектах, которые не исследованы глубоко. Так, например, очевиден успех колеса, которого не существует в природе,

или самолета, не машущего крыльями, подражая птице. К тому же пограничные науки о человеке не смогли внести существенного теоретического вклада, объясняющего хотя бы приблизительно, как протекают интеллектуальные процессы у человека, как устроена память и как человек познает окружающий мир.

Данное направление искусственного интеллекта было ориентировано на поиски алгоритмов решения интеллектуальных задач на существующих моделях компьютеров. Существенный вклад в становление новой науки внесли ее «пионеры»: Маккарти (автор первого языка программирования для задач ИИ – ЛИСПа), Минский (автор идеи фрейма и фреймовой модели представления знаний), Ньюэлл, Саймон, Шоу, Хант и другие исследователи. Отдельные аспекты истории ИИ и его ветвей хорошо освещены в работах отечественных и зарубежных классиков [Люгер, 2003; Поспелов, 1982; Стефанюк, 2011; Осипов 2022; Финн, 2020].

В 1956–1963 гг. велись интенсивные поиски моделей и алгоритмов человеческого мышления и разработка первых программ на их основе. Представители существующих гуманитарных наук – философы, психологи, лингвисты – ни тогда, ни сейчас не могли предложить таких алгоритмов. Тогда кибернетики начали создавать собственные модели. Последовательно были созданы и опробованы различные подходы:

В конце 50-х гг. родилась модель лабиринтного поиска. Этот подход представляет задачу как некоторое пространство состояний в форме графа, и в этом графе проводится поиск оптимального пути от входных данных к результирующим. В первых учебниках по искусственному интеллекту [Хант, 1978; Эндрю, 1985] описаны эти программы – они играют в игру «15», собирают «Ханойскую башню», играют в шашки и шахматы. Была проделана большая работа, но в решении практических задач идея лабиринта распространения не получила. Позднее эти алгоритмы были дополнены эвристиками.

В начале 60-х гг. – эвристикой назвали правило, теоретически не обоснованное, но позволяющее сократить количество переборов в пространстве поиска [Александров, 1975].

В 1963–1970-х гг. к решению задач стали подключать методы математической логики. Робинсон разработал метод резолюций, который позволяет автоматически доказывать теоремы при наличии набора исходных аксиом [Robinson, 1965]. Примерно в это же время выдающийся отечественный математик С. Ю. Маслов предложил так называемый обратный вывод, впоследствии названный его именем, решающий аналогичную задачу другим способом [Маслов, 1964]. На основе метода резолюций Альбер Кольмероз в 1973 г. создает язык логического программирования Пролог. Большой резонанс имела программа «Логик-теоретик», созданная

Ньюэлом, Саймоном и Шоу, которая доказывала школьные теоремы [Норвиг, Рассел, 2015]. Однако большинство реальных ситуаций не сводится к набору аксиом, и человек, решая производственные задачи, не использует классическую логику, поэтому логические модели имеют существенные ограничения по классам решаемых задач.

История искусственного интеллекта полна драматических событий, одним из которых стал в 1973 г. так называемый «доклад Лайтхилла», который был подготовлен в Великобритании по заказу Британского совета научных исследований. Известный математик Д. Лайтхилл, профессионально не связанный с ИИ, подготовил обзор состояния дел в области ИИ. В докладе были признаны определенные достижения в области ИИ, однако уровень их определялся как разочаровывающий, и общая оценка была отрицательная с позиций практической значимости. Этот отчет отбросил европейская исследовательская примерно на 5 лет назад, так как финансирование ИИ сократилось.

А в это же время существенный прорыв в практических приложениях искусственного интеллекта произошел в США, когда к середине 1970-х гг. на смену поискам универсального алгоритма мышления пришла идея моделировать конкретные знания специалистов-экспертов. В США появились первые коммерческие системы, основанные на знаниях, или экспертные системы (ЭС). Пришел новый подход к решению задач искусственного интеллекта – представление знаний. Созданы MYCIN и DENDRAL [Shortliffe, 1976; Buchanan, Feigenbaum, 1978; Джарратано, Райли, 2006], ставшие уже классическими, две первые экспертные системы для медицины и химии. В дальнейшем появились тысячи успешных ЭС для различных предметных областей – от сельского хозяйства до приборостроения [Хант, 1978; Эндрю, 1985; Попов, 1987; Рыбина, 2010; Башлыков, Еремеев, 1994; Гаврилова, 1984 и др.]. Хотя пик интереса к ЭС пройден, разработка систем, основанных на знаниях остается актуальной [Стюарт, Норвиг, 2016; Загоруйко, Загоруйко, 2022].

В ответ на успехи США в конце 1970-х гг. в гонку включается Япония, объявив о начале проекта машин V поколения, основанных на знаниях. Проект был рассчитан на 10 лет и объединял лучших молодых специалистов (в возрасте до 35 лет) крупнейших японских компьютерных корпораций. В итоге они создали достаточно громоздкий и дорогой символьный процессор, аппаратно-реализующий ПРОЛОГО подобный язык. Однако, положительный эффект этого проекта очевиден. В Японии появилась значительная группа высококвалифицированных специалистов в области ИИ, которая добилась существенных результатов в различных прикладных задачах, параллельно создав в одной из лабораторий первую нейроплату, аппаратно реализовав нейросеть.

Начиная с середины 1980-х гг. парадигма работы со знаниями получает коммерческий успех. Создаются промышленные экспертные системы, растет интерес к самообучающимся системам. Издаются десятки научных журналов, ежегодно собираются международные и национальные конференции по различным направлениям ИИ. Появляется новая наука – инженерия знаний [Adeli, 1994; Harmelen et al., 2004], исследующая модели и методы извлечения и структурирования знаний [Гаврилова, Хорошевский, 2001; Загоруйко & Загоруйко, 2020].

Далее будут рассмотрены трудности и противоречия каждого из подходов на современном этапе.

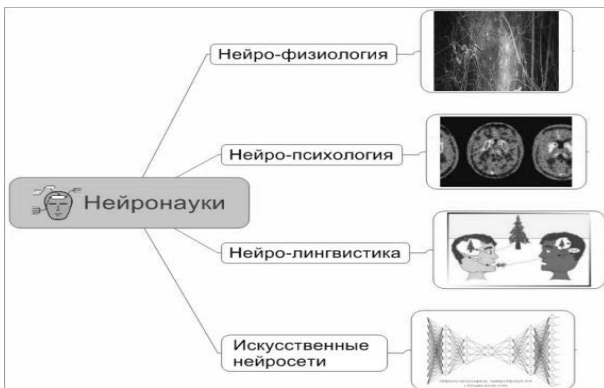
## **2. Нейросети – как статистический подход**

Бесспорные успехи нейроподхода и машинного обучения [Miikkulainen, 2018; Николенко, 2018; Миркин БГ, Кузнецов СО. Машинное обучение. Конспект лекций. – 2009.] последние 5-6 лет в области визуального распознавания образов и обработки больших данных (финансовое моделирование, маркетинговая аналитика, кибербезопасность и др.) производят сильное впечатление, и практически не оставляют места критическому анализу [Mirajilili, 2019]. Успешные алгоритмы генерации текстов и картин на основе тех же идей порождают ложные иллюзии у обывателей. При этом принцип действия любой искусственной нейронной сети остается неизменным:

“Искусственный нейрон – это такая функция, которая преобразует несколько входных сигналов (фактов, параметров ...) в один выходной на основе некой формулы порога возбуждения (threshold)”. Настройкой весов этих фактов, а также порога возбуждения – настраивается адекватность нейрона. Современные нейросети оперируют миллионами параметров и тысячами слоев, но принцип все тот же. В то время как биология и нейронауки (рис. 2) видят гораздо более сложные принципы, которые пока не поняты ими.

На основе нейросетей улучшается машинный перевод, при этом, ни на йоту не приближаясь к пониманию текста. А мыслящие аналитики хотят задать только один вопрос «Все эти алгоритмы демонстрируют способность искать сходство на сверхбольших объемах данных... и находить его, обучаясь, генерировать подобные. Где здесь «интеллектуальность?»».

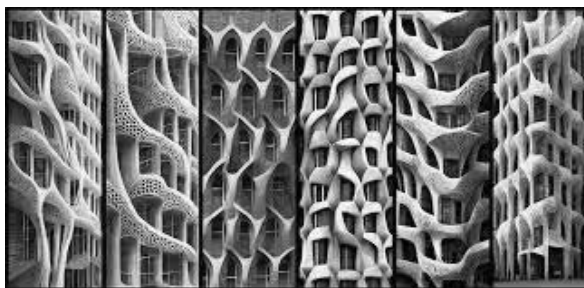
Однако, технические успехи монстров, например, алгоритмы-трансформеры GPT3 (Generative Pre-trained Transformer) от компании OpenAI и подобные им, создают внешне осмысленные тексты, например, на заданную тему, или продолжая предложения [Floridi & Chiriatti, 2020]. Правда, за нелепые ошибки недоброжелатели их называют «Генераторы умно звучащего бреда» [Рождественская, 2022].



*Рис. 2. Нейронауки*

Не утихает хайп и вокруг машинного «творчества». Условно-бесплатная платформа midjourney (Midjourney.com) конкурирующая с программой DALL-E от Open-AI, позволяет генерировать изображения, ролики в 3D или архитектурные эскизы исключительно высокого качества [Oppenlaender, 2022]. Эта платформа копирует манеры и стили определенных художников или направлений (см. рис. 3 [https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcSQhZIFJUKzjeFrgfY6olAL&dQkF6\\_7k7sQMw&usqp=CAU](https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcSQhZIFJUKzjeFrgfY6olAL&dQkF6_7k7sQMw&usqp=CAU)). На входе достаточно ввести текст-подсказку с описанием изображения (например, «Зверь с головой греческого бога побеждает современный танк в стиле Клода Моне»).

Принцип работы – все те же классические статистические пермутации (или перестановки элементов), пропущенные через миллионные обучающие выборки на примерах. Торжество данных и математики над смыслом и творчеством.



*Рис. 3. Пример архитектурного эскиза генератора изображений Midjourney*

В данном случае, правомернее говорить об искусственном “ремесле”.

По-прежнему «трудные» для людей задачи, вроде сложных расчетов, анализ финансовых рынков, подстрочные языковые переводы – просты для компьютера. «Простые» же задачи, такие как зрение, движение, восприятие – безумно трудны для него. Или, как говорил ученый, специалист по компьютерам, Дональд Кнут, «Искусственный интеллект преуспел сегодня во всем, что требует «мышления», но провалился в задачах, которые люди и животные делают, не задумываясь».

Общеизвестны ошибки и конфузы этих алгоритмов [Raji, et al., 2022], например, было обнаружено более чем 20 000 случаев мошенничества с пособиями по безработице, ложно помеченных алгоритмом MIDAS в Мичигане [Charette, 2018], и американцу Брайану Расселу пришлось объявить себя банкротом, подорвав его способность обеспечивать двух своих маленьких детей. Государство сняло с него ложные обвинения лишь два года спустя.

### **3. Работа со знаниями - в поисках смысла**

В логико-философской традиции понятие смысла чаще всего идентично понятию значения. Что означает, например, повышение температуры больного? Для понимания смыслов человек использует свои знания. Символьное представление знаний используется в интеллектуальных системах с 1976 года с момента появления баз знаний экспертных систем. Первыми моделями представления знаний выступали продукции, фреймы, семантические сети и т. п. [Harmelen et al., 2004; Загоруйко & Загоруйко, 2020]. В 1980-х годах семантические сети и фреймы были стандартизированы термином «онтология» [Gruber, 1975], которая определяется как спецификация концептуализации в виде набора понятий и отношений. Онтология – это способ формализации знаний, реализованный на основе формального описания объектов, фактов и отношений между ними. Графы знаний используют онтологии как основу представления данных, включающего терминологический словарь и набор утверждений о моделируемых объектах [Муромцев и др., 2019].

Из краткой истории ИИ, изложенной выше, следует, что зарождение символьного ИИ следует связать с появлением экспертных систем и разработке моделей представления знаний (семантические сети, фреймы и продукции). При этом традиционно акцент в исследованиях ставился на формализмы описания знаний. Однако трудности по созданию и структурированию смысловых семантических верхне-уровневых моделей предметных областей большинство исследователей обходили стороной.



Разработка этих моделей – междисциплинарная задача, требующая погружения в предметную область, ее специфику, терминологию и секреты. Задача составления простого упорядоченного словаря понятий и связей между ними является нетривиальным упражнением в области системного анализа, лингвистики и структурирования знаний конкретной предметной области [Смирнов, 2014; Добров и др., 2015]. Вот такие упорядоченные словари стали называть онтологиями, и сейчас это одна из перспективных моделей. Следует отметить, что в настоящее время помимо “ручных” активно используются автоматические способы разработки онтологий [Бова и др., 2014], что позволяет создавать онтологии в промышленных масштабах, однако, качество и семантическая глубина их иногда уступает экспертным онтологиям.

Инженерия знаний развивалась и выросла вместе с появлением новых моделей, у нее появились дочерние ветви, например *ontology engineering* (онтологический инжиниринг) [Sure, Staab, and Studer, 2009; Боргест, 2013; Смирнов, 2014; Гаврилова, 2005].

Однако, эпоха больших данных повлияла и на модели знаний. Так появились сложные многосвязные структуры, облегчающие поиск и связь данных – графы знаний. Первый граф знаний (Knowledge Graph) был создан в 2000-х компанией Google как база знаний для поисковой системы [Sheth et al, 2001], основанная на онтологиях. Существует ряд других способов представления знаний, но графы знаний стали самым популярным форматом для репрезентации знаний в конкретной области [Бхатт и др., 2020]. Графы знаний используются также в современных цифровых помощниках, например, в Amazon Alexa или Apple Siri, и ряде других систем.

Граф знаний представляет набор концептов (объектов, понятий), соединенных друг с другом отношениями. Граф может отображать знания из нескольких областей (как в базах DBpedia и Yago), знания из конкретной области (так Bio2RDF и UMLS содержат знания из области биологии и медицины) или знания, относящиеся к компании [Noy, et al., 2017].

Графы знаний основаны на смысловых или семантических связях, и ближе к интуитивному пониманию интеллектуальной деятельности.

## **Заключение. Вместе и врозь?**

Доклад посвящен размышлениям о двух ветвях ИИ – статистического, основанного на машинном обучении и больших данных, и символического, основанного на знаниях.

Пока обе ветви ИИ живут независимо, со своими успехами и провалами борются самостоятельно. На сегодняшний день громче звучит голос машинного обучения, механистического, сильного и беспощадного. Пока – это дуэль. Сложится ли дуэт покажет день грядущий.

## Список литературы

- [Adeli,1994] Adeli H. Knowledge Engineering. – McGraw-Hill Publishing Company: New York, 1994.
- [Floridi & Chiriatti, 2020] Floridi L., & Chiriatti GPT-3: Its nature, scope, limits, and consequences // *Minds and Machines*. – 30 (4). – P. 681-694.
- [Floridi, et al., 2018] Floridi L., Cowls J., Beltrametti M., Chatila R., Chazerand P., Dignum V., & Vayena. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations // *Minds and machines*. – 28 (4). – P. 689-707.
- [Harmelen & Antiniou, 2004] Harmelen F.V. & Antoniou, G., & Web ontology language: Owl. In *Handbook on ontologies*. – Springer, Berlin, Heidelberg, 2004. – P. 67-92.
- [Miikkulainen, 2019] Miikkulainen R. et al. Evolving deep neural networks // *Artificial intelligence in the age of neural networks and brain computing*. – Academic Press, 2019. – P. 293-312.
- [Mirjalili, 2019] Mirjalili S. Evolutionary algorithms and neural networks // *In Studies in computational intelligence*. Vol. 780. – Berlin/Heidelberg, Germany: Springer, 2019.
- [Oppenlaender, 2022] Oppenlaender J. The Creativity of Text-based Generative Art // arXiv preprint arXiv:2206.02904, 2022.
- [Raji et al., 2022] Raji et al. Анализ ошибок и провалов. – 2022.
- [Raji, Kumar, Horowitz & Selbst, 2022] Raji I., Kumar E., Horowitz A. & Selbst A. The Fallacy of AI Functionality. In 2022 ACM Conference on Fairness, Accountability, and Transparency, 2022. – P. 959-972.
- [Charette, 2018] Charette R. Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold-IEEE Spectrum // *IEEE Spectrum* 18, 3 (2018), 6.
- [Robinson, 1965] Robinson J. A machine-oriented logic based on the resolution principle // *Journal of the ACM (JACM)*. – 1965. – 12.1. – P. 23-41.
- [Shortliffe, Sharon, 1976] Shortliffe W., Sharon M., et al. Computerized consultation system for selection of antimicrobial therapy // *American Journal of Hospital Pharmacy*. – 1976. – 33.12. – P. 1304-1308.
- [Stottler, 2002] Stottler D. History of Artificial Intelligence. – 2002.
- [Sure, Staab, Studer, 2009] Sure Y., Steffen Staab S., Studer R. Ontology engineering methodology // *Handbook on ontologies*. – Springer, Berlin, Heidelberg, 2009. – P. 135-152.
- [Zhang & Lu, 2021] Study on artificial intelligence: The state of the art and future prospects // *Journal of Industrial Information Integration*. – 23. – 100224.
- [Александров, 1975] Александров Е.А. 1975. Основы теории эвристических решений. – М.: Советское радио.
- [Бова и др., 2014] Бова В.В., Лещанов Д.В., Кравченко Д.Ю., & Новиков А.А. Компьютерная онтология: задачи и методология построения // *Информатика, вычислительная техника и инженерное образование*. – 2014. – № 4.

- [Гаврилова, 1984] Гаврилова Т.А. Представление знаний в экспертной диагностической системе АВТАНТЕСТ // Известия АН СССР. Техническая кибернетика. – 1984. – № 5. – С. 168-175.
- [Гаврилова, 2005] Гаврилова Т.А. Об одном подходе к онтологическому инжинирингу // Новости искусственного интеллекта. – 2005. – № 3. – С. 25-31.
- [Гаврилова, Хорошевский, 2000] Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем: учебник. – СПб.: Изд-во «Питер», 2000.
- [Горбань, 2000] Нейроинформатика: кто мы, куда мы идем, как путь наш измерить // Информационные технологии. – М.: Изд-во "Машиностроение", 2000. – № 4. – С. 10-14.
- [Джонс, 2019] Джонс М.Т. Программирование искусственного интеллекта в приложениях. – М., 2019.
- [Добров, Иванов, Лукашевич, Соловьев, 2015] Онтологии и тезаурусы. – Казань: КГУ, 2015.
- [Еремеев, Башлыков, 1994] Башлыков А., Еремеев А.П., Экспертные системы поддержки принятия решений в энергетике / под. ред. А.Ф. Дьякова. – М.: Изд-во МЭИ, 1994.
- [Захаров, Попов, Пospelов, Хорошевский, 1990] Захаров В.Н., Попов Э.В., Пospelов Д.А., & Хорошевский В.Ф. Искусственный интеллект // Справочник в трех томах. М.
- [Кобринский, 2022] Кобринский, Б. А., Горячие ступени в будущее искусственного интеллекта // В «Шаг в будущее: искусственный интеллект и цифровая экономика. Технологическое лидерство: взгляд за горизонт?». – 2022. – С. 120-128.
- [Люгер, 1977] Люгер Д.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем. – 4-е изд. – Изд. дом «Вильямс», 2003.
- [Маслов, 1964] Маслов, С. Ю. Обратный метод установления выводимости в классическом исчислении предикатов // Доклады Академии наук. – 159, 1. Российская академия наук, 1964. – С. 17-20.
- [Миркин, Кузнецов, 2009] Миркин Б.Г., Кузнецов С.О. Машинное обучение // Конспект лекций. ВШЭ. – 2009.
- [Николенко, Кадури, Архангельская, 2018] Глубокое обучение. – СПб: Питер, 2018.
- [Оселедец, 2022] Оселедец И.В. Успехи и проблемы машинного обучения // Проектирование будущего // Проблемы цифровой реальности. – 5. – С. 102-108.
- [Осипов, 2022] Осипов Г. Методы искусственного интеллекта. – 2022.
- [Рыбина, 2010] Рыбина Г.В. Основы построения интеллектуальных систем. – М., 2010.
- [Стюарт, Норвиг, 2016] Стюарт Р., Норвиг С. Искусственный интеллект. Современный подход. – СПб.: Диалектика/Вильямс, 2016.
- [Пospelов, 1982] Пospelов Д.А. На пути к искусственному интеллекту. – М.: Наука, 1982.
- [Рождественская, 2020] Рождественская Я. Генератор умно звучащего бреда // Коммерсант. 14/08/2020. – <https://www.kommersant.ru/doc/4457086>.

- [Смирнов, 2014] Смирнов А.В. Онтологический подход к организации взаимодействия сервисов интеллектуального пространства при управлении гибридными системами // Искусственный интеллект и принятие решений. – 2014. – 4. – С. 42-51.
- [Смирнов, 2013] Смирнов С.В. Онтологии как смысловые модели // Онтология проектирования. – 2013. – 2 (8). – С. 12-19.
- [Стефанюк, 2011] Стефанюк В.Л. Творческое решение задач // Искусственный интеллект и принятие решений. – 2011. – 2. – С. 3-11.
- [Эндрю, 1985] Эндрю А. – М.: Мир, 1985.
- [Финн, 2020] Финн В.К. Точная эпистемология и искусственный интеллект // Научно-техническая информация. Серия 2: Информационные процессы и системы. – 2020. – 6. – С. 1-36.
- [Хант, 1978] Хант Э. Искусственный интеллект. – 1978.
- [AI, 2019] Artificial Intelligence Tutorial |AI Tutorial // Сайт Tutorial and example (<https://www.tutorialandexample.com/artificial-intelligence-tutorial>) 2019-07-01 06:51:49.

УДК 004.912, 004.853

## АВТОМАТИЧЕСКИЕ МЕТОДЫ ИЗВЛЕЧЕНИЯ ТАКСОНОМИЧЕСКИХ ОТНОШЕНИЙ ИЗ ТЕКСТОВ\*

Н.В. Лукашевич (*louk\_nat@mail.ru*)

МГУ имени М.В. Ломоносова,

Исследовательский центр доверенного искусственного интеллекта  
Института системного программирования  
имени В.П. Иванникова РАН, Москва

В статье представлен обзор подходов к автоматическому извлечению таксономических отношений (IS\_A отношений) из текстов, включая классические методы на основе лексико-семантических шаблонов, векторных представлений слов и их современное развитие. Также описаны новые подходы, основанные на языковых моделях типа BERT и использующие новые варианты работы с лексико-семантическими шаблонами, в которых модель должна заполнить нужную позицию в шаблоне. В статье также представлены различные способы тестирования качества извлечения таксономических отношений, приведены результаты этих тестирований. По достигаемым в настоящее время результатам можно сделать вывод, что, несмотря на прогресс в развитии методов извлечения таксономических отношений, качество извлечения является недостаточным для полностью автоматического построения или пополнения таксономий.

**Ключевые слова:** таксономии, отношения класс-подкласс, гипоним-гипероним, извлечение знаний из текстов.

### Введение

Тексты на естественном языке содержат большие объемы знаний о мире, о различных предметных областях, использовании разных слов и выражений в языке, что дает возможность автоматического или автоматизированного построения онтологий из текстов [Gómez-Pérez et al., 2004; Al-Aswadi et al., 2020].

---

\* Работа выполнена при поддержке гранта для исследовательских центров в области искусственного интеллекта, предоставленного Аналитическим центром Правительства РФ в соответствии с соглашением (идентификатор соглашения 000000D730321P5Q0002) и соглашением с Институтом системного программирования имени Иванникова Российской Академии наук от 2 ноября 2021 г. № 70-2021-00142.

Таксономические отношения являются наиболее известным типом отношений, востребованным при описании многих предметных областей. В разных сферах знаний эти отношения могут иметь разные названия такие, как класс-подкласс (разработка онтологий), родовидовые отношения (информационно-поисковые тезаурусы), гипоним-гипероним (лексическая семантика, тезаурус WordNet [Miller, 1995]), is\_a – отношения (искусственный интеллект).

Таксономические отношения образуют структурное ядро описания предметной области. Поэтому важной задачей является автоматизация построения таксономий, а также возможность автоматического приписывания этого отношения новым сущностям, появляющимся в текстах предметной области, так называемое пополнение таксономий (онтологий) на текстах предметной области.

В данном обзоре будут рассмотрены подходы к извлечению таксономических отношений из текстов, существующие наборы данных и подходы к оценке качества извлечения отношений, а также достигаемые на текущий момент результаты.

## **1. Методы, применяемые для извлечения таксономических отношений из текстов**

Автоматическое построение таксономий по текстам в течение уже почти 30 лет привлекает пристальное внимание исследователей [Hearst, 1992]. При построении семантических отношений по текстовым корпусам выделяют два основных подхода: лингвистический подход на основе лексико-семантических шаблонов, и подход, основанный на дистрибутивной семантике. В настоящее время также большое развитие получили комбинированные подходы. Кроме того, в связи с появлением предобученных на больших корпусах нейросетевых моделей типа BERT [Devlin et al., 2019] развиваются подходы на основе автоматического заполнения шаблонов нужного типа. Далее в тексте мы будем использовать для обозначения слова с широким значением термин *гипероним*, а для нижестоящего слов с более узким значением – термин *гипоним*.

### **1.1. Методы на основе лингвистических подходов**

**Лингвистический подход** представляет собой извлечение отношений между словами на основе их совместных вхождений в некоторый набор шаблонов. Методы, использующие шаблоны, основаны на том, что если обнаруживается некоторый заданный контекст между сущностями X и Y, то, значит, между этими сущностями существует заданное отношение. Для таксономического отношения таким контекстом может быть контекст «X – это вид Y». Такой подход обладает высокой точностью и простотой, может

быть применен к различным типам отношений. Работы по извлечению таксономических отношений по шаблонам начались с работы [Hearst, 1992], в которой было предложено несколько наиболее частотных шаблонов.

Применение шаблонов для извлечения отношений обычно имеет высокую точность, но также связано с несколькими проблемами, снижающими полноту извлечения отношений.

Во-первых, в используемой текстовой коллекции может не встретиться ни одного явного упоминания контекстов нужного типа, несмотря на то, что X и Y находятся между собой в нужном отношении. Из-за этого подход обладает очень низкой полнотой. Во-вторых, для качественного извлечения отношений нужно задать большое количество шаблонов.

Для увеличения полноты лексико-семантических шаблонов было предложено несколько методов. В ряде работ предлагалось использовать расширенный набор лексико-семантических шаблонов. Например, авторы работы [Seitner et al., 2016] используют набор из 59 шаблонов, собранных из разных работ.

В работе [Roller et al., 2018] для улучшения качества предсказания для низкочастотных слов, создается матрица взаимной встречаемости слов в заданных лексико-семантических шаблонах, оцененной с помощью формулы взаимной информации. К получившейся матрице затем применяется SVD-разложение и снижение размерности, так называемый метод латентного семантического анализа [Deerwester S. et al., 1990]. В результате для слов появившихся в схожих шаблонах с похожим множеством слов предсказываются одинаковые гиперонимы, что ведет к улучшению качества предсказания гиперонимов для низкочастотных слов или слов, редко упоминавшихся в используемых шаблонах.

Еще одним подходом к увеличению полноты покрытия шаблонами является так называемый бутстраппинг (частичное обучение с учителем), то есть несколько шаблонов извлечения заданного отношения задаются вручную, а извлекаемые на основе этих шаблонов примеры отношения используются для сбора новых видов шаблонов [Tian et al., 2012]. Однако при нарастании числа шаблонов точность извлечения отношений обычно снижается.

При создании шаблонов можно заметить, что некоторые из них имеют схожий вид. В частности, система Patty [Nakashole et al., 2012] увеличивает полноту извлечения отношений шаблонами за счет обобщения шаблонов. Например, шаблоны «X inc. is a Y», «X group is a Y», «X organization is a Y» можно обобщить до «X \* is a Y», что однако может привести к некоторому снижению точности извлечения отношения.

Наконец, в лексико-семантические шаблоны могут вклиниваться лишние слова, поэтому в ряде работ предлагается использовать для извлечения шаблонов синтаксические структуры предложений, а сам шаблон за-

писывать на языке синтаксических отношений [Snow et al., 2004; Aldine et al., 2018], которые могут соединять и удаленно расположенные слова синтаксического шаблона.

Для русского языка лексико-семантические шаблоны для извлечения родовидовых отношений, отношений гипоним-гипероним рассматриваются в работах [Bolshakova et al., 2015; Sabirova et al., 2014]. Авторы работы [Kiselev et al., 2015] применяли лексико-семантические шаблоны для извлечения гиперонимов слов из их толкований в толковых словарях.

## 1.2. Методы на основе дистрибутивной семантики

Второе основное направление извлечения таксономических отношений основано на методах дистрибутивной семантики и на так называемой дистрибутивной гипотезе, которая заключается в том, что семантическое сходство между словами (или другими языковыми единицами) может моделироваться на основе сравнения (сходства) контекстов этих слов [Harris 1954; Lenci 2008], т.е. предполагается, что сходство контекстов слов коррелирует с семантическим сходством слов.

Классический подход дистрибутивной семантики заключается в том, что для каждого целевого слова  $w$  вычисляется матрица, в которой каждая строка соответствует целевому слову, а каждый столбец представляет собой слова, которые встречались в контексте исходного слова. Таким образом, каждое слово представляется вектором слов, которые встречаются в его контексте, т.е. геометрически слово представляется точкой в  $n$ -мерном пространстве, размерность которого равна количеству разных слов, найденных в текстовой коллекции [Turney et al., 2010; Erk, 2012].

Задача нахождения сходства между словами сводится в таком случае к определению близости расположения соответствующих точек. Близость расположения точек может вычисляться на основе подсчета Евклидова расстояния между точками. Однако в этом случае расстояние будет сильно зависеть от частоты употребления слова в корпусе. Поэтому предполагается, что небольшой угол между векторами, исходящими из точки начала координат важнее, чем близость по Евклидову расстоянию. Для вычисления близости по углу можно считать косинусную меру близости, в которой предполагается, что чем меньше угол, тем ближе косинус этого угла к нулю. В таком случае сходство между словами можно искать с помощью скалярного произведения между векторами-контекстами [Turney et al., 2010; Erk, 2012]. Чем ближе скалярное произведение к 1, тем выше семантическое сходство между словами.

В настоящее время большую популярность получил подход к построению векторных представлений слов на основе нейронных сетей. Считают-



ся, что во многих задачах нейросетевые подходы превосходят традиционные подходы дистрибутивной семантики, кроме того, созданы более эффективные инструменты расчета этих представлений.

В 2013 году в работе [Mikolov et al., 2013] была предложена эффективная реализация метода создания векторных представлений слов относительно небольшого числа размерностей (обычно от 100 до 1000) на основе нейронных сетей – пакет word2vec. Основная идея подхода состоит в том, чтобы обучать такие векторные представления (называемые word embeddings – эмбединги) на текстовом корпусе в процессе предсказания слов в контексте для одного или нескольких слов. В настоящее время, опубликованы уже насчитанные векторные представления на корпусах большого объема, которые позволяют использовать их в различных задачах автоматической обработки текстов без дополнительных вычислительных затрат [Kutuzov et al., 2017; Wojanowski et al., 2017].

При извлечении таксономических отношений на основе векторных представлений предполагается, что гипонимы и гиперонимы, как близкие по смыслу слова, используются в похожих контекстах, что должно приводить к формированию похожих векторных представлений. Однако нужно отметить, что векторные представления слов сами по себе не позволяют определить тип отношения, которым связаны два слова. Близкими по векторным представлениям к исходному слову могут оказаться как синонимы, гипонимы или гиперонимы целевого слова, так и антонимы (*дорогой – дешевый*), а также тематически связанные слова (например, *собака – ошейник*) и др. Поэтому исходные векторные представления слов используются как основа для дальнейшего применения двух базовых подходов для извлечения гиперонимов: подходов без учителя (unsupervised) и подход с учителем (supervised).

В первой группе методов без учителя подразумевается, что существует некоторое преобразование исходных векторов слов, которое лучше соответствует отношению гипоним-гипероним. В частности, может предполагаться, что контексты слова-гиперонима включают контексты слова гипонима – так называемая гипотеза дистрибутивного включения (Distributional Inclusion Hypothesis) [Roller et al., 2014].

Вторая группа методов (с учителем) использует исходные векторные представления как признаки для последующего применения специализированных классификаторов, обученных на наборах данных гипоним-гипероним [Baroni et al., 2012]. Первые подходы с учителем использовали арифметические операции над векторами слов (сложение, вычитание, конкатенация), к которым затем применялся классификатор (логистическая регрессия, метод опорных векторов и т.п.) для предсказания гиперонима для слова [Baroni et al., 2012; Roller et al., 2014; Shwartz et al., 2016].

В рамках дальнейшего развития методов с учителем на основе векторного пространства эмбедингов было сделано предположение, что можно попытаться найти линейное преобразование, которое преобразует гипоним в гипероним – данный подход называется *projection learning* [Fu et al., 2014]. Однако в дальнейших экспериментах было показано, что невозможно найти единое преобразование гипонима в гипероним, поскольку оптимальное преобразование отличается для разных семантических классов слов. Поэтому были предложены методы одновременной семантической кластеризации слов и поиска подходящего преобразования гипоним-гипероним, соответствующего каждому семантическому классу [Yamane et al., 2016; Ustalov et al., 2017; Bernier-Colborne et al., 2018].

Для вышеупомянутых операций в векторном пространстве слов важен исходный корпус, в частности его размер и жанр текстов (интернет-страницы, Википедия, тексты предметной области и др.). Для улучшения качества векторных представлений был предложен подход на основе так называемых мета-эмбедингов, т.е. объединений различных векторных представлений, полученных на разных текстовых корпусах [Bollegala et al., 2018, Tikhomirov et al., 2021]. В качестве мета-эмбедингов могут выступать такие комбинации векторов как конкатенация или усреднение. Более гибким методом является подход на основе так называемых автокодировщиков, задача которых, стартуя с нескольких различных векторных представлений, найти наилучшее комбинированное представление (мета-эмбединг), из которого можно наилучшим образом восстановить исходные представления, а также добиться наилучшего соответствия мета-эмбедингов некоторому дополнительному критерию. В задаче предсказания гиперонимов таким критерием может быть использование так называемого триплет-лосса – мета эмбединг должен быть похож на вектора близких по смыслу слов и не похож на вектора случайных слов [Tikhomirov et al., 2021].

### **1.3. Комбинированные методы извлечения таксономических отношений**

Лингвистические методы извлечения отношений могут использовать лексико-семантические шаблоны или синтаксические отношения слов [Snow et al., 2004], совместную встречаемость слов в предложении (Yang, Callan 2009) и др. Методы извлечения отношений на основе дистрибутивных подходов базируются на извлечении контекстов употребления слов и вычислении их сходства как векторов, т.е. эти подходы позволяют извлекать отношения, которые не указываются в текстах явным образом.

Достоинства обоих методов сочетают методы инкрементального пополнения таксономии [Snow et al. 2006; Yang et al., 2009], при которых все вышеперечисленные характеристики (встречаемость в шаблонах, наличие

сходства в разного вида контекстах (линейных и синтаксических, локальных и глобальных), совместная встречаемость) служат как признаки, на основе которых принимается решение о присоединении очередного слова к создаваемой таксономии. При принятии решений о присоединении к таксономии в работе [Snow et al., 2006] используется принцип максимизации условной вероятности отношения на основе имеющихся признаков, в работе [Yang, et al., 2009] используется принцип оптимизации таксономической структуры и моделировании уровня абстрактности слова в таксономической структуре.

В модели HypeNET [Shwartz et al., 2016] для представления предложения используются три вектора: два вектора исходных слов и третий вектор, который представляет собой векторное представление для синтаксического пути (=пути в синтаксическом дереве) между двумя словами. Для создания вектора синтаксического пути используется рекуррентная нейронная сеть LSTM. Результаты данной работы значительно превосходили предыдущие результаты в задаче извлечения гиперонимов как задачи классификации, по сравнению с лингвистическими и дистрибутивными подходами, поскольку модель еще обучается еще и оптимальному кодированию синтаксических путей между словами вместо заранее заданных лексико-семантических шаблонов.

#### **1.4. Предсказание гиперонимов посредством порождения ответов на вопрос**

В связи с появлением больших предобученных нейросетевых моделей типа BERT [Devlin et al., 2019] произошло улучшение в решении многих задач автоматической обработки текстов. В этой модели было предложено использовать нейронную сеть архитектуры трансформер с механизмом самовнимания [Vaswani et al., 2017] для предобучения на больших объемах неразмеченных текстовых данных для решения двух задач: 1) предсказания маскированного слова в контексте, 2) предсказания, следует ли второе предложение из первого. В результате модель BERT создает контекстуализированные (зависящие от контекста) представления слов, которые позволяют лучше учесть контекст и решать многие задачи автоматической обработки текстов.

Специфичный подход к обучению модели BERT дал дополнительные возможности решения задач автоматической обработки текстов на основе специально заданного вопроса (подсказки или prompt) [Schick and Schutze, 2021]. Такой подход стал применяться и в задаче предсказания гиперонимов [Xu et al., 2022; Peng et al., 2022]. В частности, в методе TaxoPrompt [Xu et al., 2022] в качестве подсказки используется следующий шаблон: “What is parent-of [X]? It is [MASK] (Что есть родитель [X]? Это [MASK]).

В слот X вставляется новое слово, для которого нужно определить гипероним. Далее данный шаблон подается на вход языковой модели BERT, особенность обучения которой является то, что модель обучается предсказывать слово, скрытое маской [Devlin et al., 2019]. В качестве альтернативных подсказок могут использоваться также следующие: 1) [X], [MASK]. 2) [X] is a [MASK]. 3) [MASK], such as [X].

Для придания дополнительного контекста на вход языковой модели описанный лексико-семантический шаблон подается вместе со вторым предложением, отделенным служебным токеном [SEP]. Вторым предложением может быть толкование целевого слова из словаря, или пути их базовой онтологии, записанные в виде текстовых последовательностей. Для обучения модели строится обучающая коллекция с правильными и неправильными ответами из какой-либо таксономии. Таким образом, модель, предсказывая гипероним, использует и громадный объем данных, на котором происходило предобучение, и толкования из словаря, и векторное представление таксономии.

## **2. Тестирование методов извлечения таксономических отношений**

Для оценки качества извлечения таксономических отношений используются различные подходы, которые могут быть разделены на четыре основных группы.

**1. Предсказание отношения гиперонимии как задача классификации на заданном датасете.** При такой постановке тестирование подходов осуществляется на основе специального набора данных. Набор содержит позитивные примеры, между которыми имеется нужное отношение, и негативные примеры, между которыми отношение отсутствует. Позитивные примеры обычно берутся из некоторых известных онтологических ресурсов, отрицательные примеры порождаются в результате некоторой процедуры. Таким образом, задача тестирования ставится как задача классификации пары слов из датасета на два класса: есть нужное отношение между словами или нет. Оценка качества в этой задаче производится традиционными мерами для классификации: точность, полнота, F-мера.

В работе [Schwartz et al., 2017] представлен обзор различных операций над векторами слов, которые использовались в задачах идентификации гиперонимов и представлено тестирование предсказания гиперонимов на нескольких наборах данных в режимах обучения без учителя (unsupervised) и с учителем (supervised). Показано, что на нескольких наборах данных методы предсказания гиперонимов без учителя превышают качество 0.85, а с учителем на некоторых данных получают идеально верные ответы.

На основе в таких тестированиях результатов можно предположить, что задача определения гиперонима для слова решена. Однако она «решена» только в рамках созданных датасетов. Проблемой такого подхода к тестированию извлечения гиперонимов является то, что данная постановка не является практической. Обычно в таких наборах данных количество позитивных и негативных примеров сравнимо между собой, а на практике количество таксономических отношений между словами на порядки меньше, чем потенциальные отношения между разными парами слов в текстовом корпусе. Например, одним из известных и больших наборов, применяемых для тестирования извлечения отношений гипоним-гипероним, является датасет BLESS [Baroni et al., 2012]. BLESS содержит гиперонимы для 200 конкретных, в основном однозначных существительных. Отрицательные пары содержат ко-гипонимы, меронимы (слова, находящиеся в отношении часть-целое) и случайные пары слов, в итоге датасет содержит всего 14542 пары слов с 1337 положительными примерами, т.е. положительных пар лишь в 10 раз меньше, чем отрицательных, что не соответствует ситуации при извлечении из реальных текстовых коллекций

Кроме того, ряд исследований показали, что часто применяемые методы классификации отношений не обучаются классифицировать пару слов, а скорее запоминают типовую роль слова как гипонима или гиперонима. Например, есть слова, такие как *животное*, *растение*, *здание*, которые являются типичными гиперонимами и могут неоднократно в таком качестве использоваться в датасетах [Levy et al., 2015; Schwartz et al., 2016]. Таким образом, результаты методов могут завышаться - так называемая проблема лексического запоминания (lexical memorization) [Levy et al., 2015].

**2. Извлечение гиперонимов из заданного корпуса.** При такой постановке тестирования системы должны правильно извлечь гиперонимы из корпуса для заданного набора слов, т.е. системы должны сами найти подходящие кандидаты в гиперонимы. На тестировании SemEval-2018 [Camacho-Collados et al., 2018] эта задача ставится как задача ранжирования, т.е. автоматическая система должна создать упорядоченный список кандидатов в гиперонимы для целевых слов, причем правильные кандидаты должны находиться в начале списка. Для оценки качества подходов в этой задаче применяются такие меры как MRR – средняя обратная величина первого правильного ответа, а также MAP (Mean Average Precision), которая представляет собой комбинированную меру точности и полноты и вычисляется как сумма точностей в момент поступления в список правильного ответа, усредненную на количество правильных ответов.

В качестве базовых методов для сравнения в тестировании SemEval-2018 применялись методы предсказания гиперонимов на основе дистрибутивных представлений, которые при тестировании на

ограниченных датасетах показывали точность выше 0.7. В данной постановке задачи эти методы выдавали первый правильный гипероним слова в среднем на 30 позиции (MRR 0.03). Это связано с тем, что наиболее похожие на целевое слова по векторной модели слова часто оказываются не гиперонимами, а так называемыми ко-гипонимами (т.е. гипонимами одного и того же гиперонима).

Лучший результат в тестировании SemEval-2018 был достигнут системой CRIM [Bernier-Colborne et al., 2018], в которой был применен комбинированный подход, включающий лексико-семантические шаблоны для определения гиперонимов и ко-гипонимов, метод типа projection-learning с обучением преобразования пространства от гипонима к гиперониму с одновременным обучением семантических кластеров. Данная система достигла качества 0.33 MRR, что означает, что правильный гипероним предсказывается в среднем на 3 позиции.

В более позднем подходе [Bai et al., 2021] предлагается использовать не один переход в трансформации матрицы гипоним-гипероним, как в исходном подходе projection learning, но и использовать такую трансформацию несколько раз, что выявить вышестоящие гиперонимы как в цепочке: *Билл Клинтон – политик – человек*. Данный подход достигает MRR-0.39 на данных SemEval-2018.

**3. Автоматическое порождение таксономии** Третьим способом тестирования является задача автоматического порождения таксономии на основе текстового корпуса и сравнения порожденной таксономии с существующей. Такая задача ставилась на семинарах SemEval-2015, 2016 [Bordea et al., 2015, Bordea et al., 2016]. Таким образом, нужно не только извлечь отношения гиперонимии для некоторого множества целевых слов, но и выстроить из них правильную структуру ациклического графа. В качестве тестовых данных были использованы существующие таксономии из разных областей: химические вещества, оборудование, еда, науки, объемом от 450 вершин (науки) до 17 тысяч вершин (химические вещества). Сами вершины заданы, для них нужно определить таксономические отношения. Слова могут иметь более одного гиперонима, тем самым таксономия может быть полииерархией.

Оценка осуществлялась как стандартная задача классификации с использованием мер точности, полноты и F-меры по сравнению с эталонными ресурсами. Также были выполнены оценки структурного качества полученной таксономии (например, отсутствие циклов). Большинство участников использовали Википедию как корпус и применяли комбинированные методы, включающие лексико-семантические шаблоны и дистрибутивное сходство термов по корпусу. Максимальный результат лучшей системы достиг f-меры 0.39 (точность 0.36). Ручная проверка точности фрагмента таксономии дала лучший результат около 0.6.

**4. Автоматическое пополнение таксономии.** При такой постановке предполагается, что некоторая общая таксономия имеется, и нужно ее достроить в некоторой предметной области или на основе новых данных. Впервые решение такой задачи предлагалось и тестировалось на основе пополнения WordNet в работе [Snow et al., 2006].

Также такая задача, например, ставилась в тестировании SemEval-2016 task 14 [Jurgens et al., 2016]. Участникам тестирования нужно было найти место нового слова в иерархии WordNet [Miller et al., 1995] в качестве синонима или гипонима к существующему понятию. Особенность задачи состояла в том, чтобы достроить таксономию WordNet за счет существующих словарей, поэтому новые слова были снабжены толкованиями из этих словарей. Качество измерялось двумя мерами: Accracy – точность установления отношения с учетом расстояния (в количестве отношений) от правильного ответа и Recall – доля слов, для которых система предложила решение. Результаты лучшей системы были немного лучше, чем результаты базового решения, которое выбирало в качестве гиперонима нового слова наиболее частотное значение первого слова той же части речи, что и новое слово, в толковании. При этом точность базового метода была немногим более 0.5, ответы были даны для всех слов, что привело к значению F-меры – 0.68.

Лучшие на текущий момент результаты в тестировании SemEval-2016 task 14 получены системой TaxoExpan [Shen et al. 2020], достигшая точности 0.566 и значения F-меры 0.723. Система представляет новое слово и слова-кандидаты в WordNet как эмбединги fastText, толкование представляется как усредненный вектор эмбедингов упомянутых слов. Представление слова-кандидата как усреднение векторов токена и толкования подается в систему подбора подходящего гиперонима, основанную на представлении понятий в таксономии WordNet с использованием графовых нейронных сетей (GNN).

В 2020 году в рамках конференции Диалог было организовано открытое тестирование по пополнению таксономии для русского языка RUSSE-2020 [Nikishina и др. 2020a]. В тестировании использовались две версии русского варианта WordNet – RuWordNet [Loukachevitch et al., 2016]. Датасет для пополнения состоял из слов из неопубликованной новой версии RuWordNet (RuWordNet 2.0), участвующие системы должны были подобрать подходящий гипероним (точнее синсет – понятие в RuWordNet) в опубликованной версии RuWordNet. В качестве правильных ответов зачислялись как указанные в RuWordNet 2.0 гиперонимы, так и их вышестоящие на 1 шаг гиперонимы (гиперонимы второго порядка), таким образом моделировалась работа «интеллектуального» помощника эксперта,

в рамках которой должно быть найдено приближительное место в семантической сети для привязывания нового слова. Качество предсказания оценивалось с использованием мер средней точности и MRR.

Базовым компонентом, использованным в работах большинства участников, было извлечение наиболее похожих слов из RuWordNet для целевого слова на основе векторных представлений; синсеты, гиперонимы и гиперонимы второго порядка использовались для формирования синсетов кандидатов в качестве гиперонимов для целевого слова. Этот же подход использовался при создании базового решения, предоставленного организаторами. Многие участники отмечали, что для достаточно конкретных слов, которые входили в набор данных для тестирования, наиболее похожими словами оказались не синонимы или гиперонимы, а ко-гипонимы, т.е. гипонимы того же гиперонима.

Лучшим векторным представлением для предсказания гиперонимов оказались эмбединги fasttext [Bojanowski et al., 2017], рассчитанные на интернет-корпусе Common Crawl и Википедии<sup>1</sup>. Лучшим методом в данном тестировании оказался метод, использующий разнообразные признаки, включая кандидаты на основе эмбедингов, словарные толкования Викисловаря, а также анализ результатов поиска в поисковых системах Яндекс и Google. Данный подход достиг результата 0.55 MRR для предсказания гиперонимов существительных, что соответствует нахождению в среднем первого правильного ответа на втором месте выдачи.

В продолжение данного тестирования и следуя его основной идее (тестирование пополнения таксономии на основе разных версий одного и того же ресурса), в 2021 году был создан специализированный датасет, использующий три последовательных версии англоязычного WordNet и две версии RuWordNet - датасет был назван Diachronic wordnets [Nikishina et al., 2020b]. В датасеты входят новые слова более поздних версий, которым нужно предсказать синсет-гипероним, используя более старую версию ресурсов. Таким образом, на данном датасете можно проверять разные подходы пополнения ресурсов типа WordNet параллельно на двух языках.

На созданных наборах данных лучшие результаты были получены на основе мета-эмбедингов, которые комбинируют разные векторные представления, включая эмбединги слов, полученные разными методами на разных корпусах, векторные представления вершин пополняемых ресурсов, а также информацию из толкований Викисловаря. В частности, лучший результат на датасете RUSSE-2020 удалось повысить до 0.6 MRR без всякого использования результатов поисковых систем Интернет [Nikishina et al., 2022].

---

<sup>1</sup> <https://fasttext.cc/docs/en/crawl-vectors.html>.



## Заключение

Таксономические отношения, связывающие сущность с ее классом, являются центральным отношением при представлении знаний во многих предметных областях. Поэтому много внимания уделяется автоматическим методам выявления таксономических отношений из текстов.

В данной статье мы представили обзор подходов к автоматическому извлечению таксономических отношений между словами, начиная от самых классических таких как лексико-семантические шаблоны. В настоящее время базой для извлечения таксономических отношений являются векторные представления слов, которые отражают сходство контекстов упоминания гипонима и его гиперонима. В статье также представлены новые подходы, работающие на основе языковых моделей типа BERT, и использующие новые варианты работы с лексико-семантическими шаблонами, в которых модель заполняет позицию гиперонима.

В статье также описаны различные способы тестирования качества извлечения таксономических отношений, а также приведены результаты этих тестирований, из которых следует, что несмотря на прогресс в развитии методов извлечения таксономических отношений, качество извлечения является недостаточным для полностью автоматического построения или пополнения таксономий.

## Список литературы

- [Al-Aswadi et al., 2020] Al-Aswadi F.N., Chan H.Y., Gan K.H. Automatic ontology construction from text: a review from shallow to deep learning trend // Artificial Intelligence Review. – 2020. – Vol. 53, No. 6. – P. 3901-3928.
- [Aldine et al., 2018] Aldine A.I., Harzallah M., Berio G., Bechet N., Faour A. Redefining Hearst patterns by using dependency relations // Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. – Vol. 2: KEOD, 2018. – P. 148-155.
- [Bai et al., 2021] Bai Y., Zhang R., Kong F., Chen J., Mao Y. Hypernym Discovery via a Recurrent Mapping Model // Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. – 2021. – P. 2912-2921.
- [Baroni et al., 2012] Baroni M., Bernardi R., Do N.Q., Shan C.C. Entailment above the word level in distributional semantics // Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. – 2012. – P. 23-32.
- [Bojanowski et al., 2017] Bojanowski P., Grave E., Joulin A., Mikolov T. Enriching word vectors with subword information // Transactions of the Association for Computational Linguistics. – 2017. – 5. – P. 135-146.
- [Bernier-Colborne et al., 2018] Bernier-Colborne G., Barri`ere C. CRIM at SemEval-2018 task 9: A hybrid approach to hypernym discovery // Proceedings of The 12th International Workshop on Semantic Evaluation. – 2018. – P. 725-731.

- [**Bollegala et al., 2018**] Bollegala D., Bao C. Learning word meta-embeddings by autoencoding // Proceedings of the 27th International Conference on Computational Linguistics. – 2018. – P. 1650-1661.
- [**Bolshakova et al., 2015**] Bolshakova E.I., Efremova N.E. A heuristic strategy for extracting terms from scientific texts // International Conference on Analysis of Images, Social Networks and Texts. – Springer, Cham, 2015. – P. 297-307.
- [**Bordea et al., 2015**] Bordea G., Buitelaar P., Faralli S., Navigli R. Semeval-2015 task 17: Taxonomy extraction evaluation (texeval) // Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 2016. – pp. 1081–1091.
- [**Bordea et al., 2016**] Bordea G., Lefever E., Buitelaar P. Semeval-2016 task 13: Taxonomy extraction evaluation (texeval-2) // Proceedings of the 10th international workshop on semantic evaluation (semeval-2016). – 2016. – P. 1081-1091.
- [**Camacho-Collados et al., 2018**] Camacho-Collados J., Delli Bovi C., Espinosa-Anke L., Oramas S., Pasini T., Santus E., Shwartz V., Navigli R., Saggion H. Semeval2018 task 9: Hypernym discovery // Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval-2018); ACL; 2018. – P. 712-724.
- [**Deerwester S. et al., 1990**] Deerwester S., Dumais S.T., Furnas G.W., Landauer T.K., & Harshman R. Indexing by latent semantic analysis // Journal of the American society for information science. – 1990. – Vol. 41, No. 6. – P. 391-407.
- [**Devlin et al., 2019**] Devlin J., Ming-Wei Chang, Lee K., Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding // Proceedings of NAACL-2019. – 2019. – P. 4171-4186.
- [**Erk, 2012**] Erk K. Vector space models of word meaning and phrase meaning: A survey // Language and Linguistics Compass. – 2012. – Vol. 6, No. 10. – P. 635-653.
- [**Fu et al., 2014**] Fu R., Guo J., Qin B., Che W., Wang H., Liu T. Learning semantic hierarchies via word embeddings // Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). – 2014. – P. 1199-1209.
- [**Gómez-Pérez et al., 2004**] Gómez-Pérez A., Manzano-Macho D. An overview of methods and tools for ontology learning from texts // The knowledge engineering review. – 2004. – Vol. 19, No. 3. – P. 187-212.
- [**Hearst., 1992**] Hearst M. Automatic acquisition of hyponyms from large text corpora // In Proceedings of the 14th conference on Computational linguistics. – 1992. – Vol. 2. – P. 539-545.
- [**Harris, 1954**] Harris Z. Distributional structure // Word. – 1954. – 10 (23). – P. 146-162.
- [**Jurgens et al., 2016**] Jurgens D., Pilehvar M. Semeval-2016 task 14: semantic taxonomy enrichment // In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). – 2016. – P. 1092-1102.
- [**Kiselev et al., 2015**] Kiselev Y., Porshnev S., Mukhin M. Method of extracting hyponym-hypernym relationships for nouns from definitions of explanatory dictionaries (in Russian) // Software Engineering. – 2015. – 10. – P. 38-48.
- [**Kutuzov et al., 2017**] Kutuzov A., Kuzmenko E. WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models / In: Ignatov D. et al. (eds) // Analysis of Images, Social Networks and Texts. AIST 2016. Communications in Computer and Information Science. – Vol. 661. – Springer, Cham, 2017.

- [Lenci, 2008] Lenci A. Distributional semantics in linguistic and cognitive research // Italian journal of linguistics. – 2008. – Vol. 20, No. 1. – P. 1-31.
- [Levy et al., 2015] Levy O., Remus S., Biemann Ch., Dagan I. Do supervised distributional methods really learn lexical inference relations? // In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. – 2015. – P. 970-976.
- [Loukachevitch et al., 2016] Loukachevitch N., Lashevich G., Gerasimova A., Ivanov V., Dobrov B. Creating Russian wordnet by conversion // In Computational Linguistics and Intellectual Technologies: papers from the Annual conference” Dialogue. – 2016. – P. 22-30.
- [Mikolov et al., 2013] Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed representations of words and phrases and their compositionality // In Advances in neural information processing systems. – 2013. – P. 3111-3119.
- [Miller, 1995] Miller G.A. WordNet: a lexical database for English // Communications of the ACM. – 1995. – Vol. 38, No. 11. – P. 39-41.
- [Nakashole et al., 2012] Nakashole N., Weikum G., Suchanek F. PATTY: A taxonomy of relational patterns with semantic types // Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. – 2012. – P. 1135-1145.
- [Nikishina et al., 2020a] Nikishina I., Logacheva, V., Panchenko, A., Loukachevitch, N. RUSSE’2020: Findings of the First Taxonomy Enrichment Task for the Russian Language // Computational Linguistics and Intellectual Technologies: papers from the Annual conference “Dialogue”. – 2020.
- [Nikishina et al., 2020b] Nikishina I., Panchenko, A., Logacheva, V., Loukachevitch, N. Studying Taxonomy Enrichment on Diachronic WordNet Versions // Proceedings of the 28th International Conference on Computational Linguistics: Association for Computational Linguistics. – 12. 2020.
- [Nikishina et al., 2022] Nikishina I., Tikhomirov, M., Logacheva, V., Nazarov, Y., Panchenko, A., & Loukachevitch, N. Taxonomy Enrichment with Text and Graph Vector Representations // Semantic Web. – 2022. – Vol. 13, No. 3. – P. 441-475.
- [Roller et al., 2014] Stephen Roller, Katrin Erk, and Gemma Boleda. Inclusive yet selective: Supervised distributional hypernymy detection // In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. – 2014. – P. 1025-1036.
- [Roller et al., 2018] Roller S., Kiela D., Nickel M. Hearst Patterns Revisited: Automatic Hypernym Detection from Large Text Corpora // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2018. – pp. 358-363.
- [Sabirova et al., 2014] Sabirova K., Lukanin A. Automatic Extraction of Hypernyms and Hyponyms from Russian Texts // Supplementary Proceedings of the 3rd International Conference on Analysis of Images, Social Networks and Texts (AIST 2014). – 2014. – P. 35-40.
- [Schick et al., 2021] Schick T., Schutze H. Exploiting cloze-questions for few-shot text classification and natural language inference // Proceedings EACL-2021. – 2021. – P. 255-269.

- [Seitner et al., 2016] Seitner J., Bizer C., Eckert K., Faralli S., Meusel R., Paulheim H., Ponzetto S.P. A large database of hypernymy relations extracted from the web // LREC-2016. – 2016.
- [Shen et al., 2020] Shen J., Shen, Z., Xiong, C., Wang, C., Wang, K., Han, J. TaxoExpan: Self-supervised taxonomy expansion with position-enhanced graph neural network // Proceedings of The Web Conference 2020. – 2020. – P. 486-497.
- [Snow et al., 2004] Snow R., Jurafsky D., Ng A. Learning syntactic patterns for automatic hypernym discovery // Advances in neural information processing systems. – 2004.
- [Snow et al., 2006] Snow R., Jurafsky D., Ng A. Semantic taxonomy induction from heterogeneous evidence // In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. – 2006. – P. 801-808.
- [Shwartz et al., 2016] Shwartz V., Goldberg Y., Dagan I. Improving hypernymy detection with an integrated path-based and distributional method // In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. – 2016. – P. 2389-2398.
- [Shwartz et al., 2017] Shwartz V., Santus E., Schlechtweg D. Hypernyms under Siege: Linguistically-motivated Artillery for Hypernymy Detection // Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Vol. 1, Long Papers, 2017. – P. 65-75.
- [Tian et al., 2012] Tian F., Yuan C., Ren F. Hyponym extraction from the web by bootstrapping // IEEJ transactions on electrical and electronic engineering. – 2012. – Vol. 7, No. 1. – P. 62-68.
- [Tikhomirov et al., 2021] Tikhomirov M., Loukachevitch N. Meta-Embeddings in Taxonomy Enrichment Task // Computational Linguistics and Intellectual Technologies: papers from the Annual conference Dialogue. – 2021. – P. 681-691.
- [Turney et al., 2010] Turney P.D., Pantel P. From frequency to meaning: Vector space models of semantics // Journal of artificial intelligence research. – 2010. – Vol. 37. – P. 141-188.
- [Ustalov et al., 2017] Ustalov D., Arefyev, N., Biemann, C., & Panchenko, A. Negative Sampling Improves Hypernymy Extraction Based on Projection Learning // Proceedings of EACL 2017. – 2017. – P. 543.
- [Vaswani et al., 2017] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., ... & Polosukhin I. Attention is all you need // Advances in neural information processing systems. – 2017. – Vol. 30.
- [Yamane et al., 2016] Yamane J., Takatani T., Yamada H., Miwa M., & Sasaki Y. Distributional hypernym generation by jointly learning clusters and projections // Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. – 2016. – P. 1871-1879.
- [Yang et al., 2009] Yang H., Callan J. A metric-based framework for automatic taxonomy induction // Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009. – P. 271-279.

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ДЛЯ КИБЕРБЕЗОПАСНОСТИ: ПРОТИВОСТОЯНИЕ ТЕХНОЛОГИЙ НАПАДЕНИЯ И ЗАЩИТЫ\*

И.В. Котенко (*ivkotel@mail.ru*)

Санкт-Петербургский Федеральный исследовательский центр РАН  
(СПб ФИЦ РАН), Санкт-Петербург

Состояние кибербезопасности в настоящее время представляет собой большую социальную проблему, и для обеспечения кибербезопасности необходимы инновационные решения. Недавние достижения в области технологий искусственного интеллекта (ИИ) привели к росту инноваций и автоматизации процессов защиты информации. ИИ стал жизнеспособным подходом к обработке огромных объемов разнородных данных и выполнения фундаментальных задач кибербезопасности, в том числе управления уязвимостями и оценки защищенности, мониторинга безопасности, приоритизации активов, распределенного контроля доступа. Хотя использование методов и технологий ИИ предполагает существенное увеличение эффективности обеспечения кибербезопасности, однако эти методы и технологии могут использоваться и злонамеренно. Целенаправленные кибератаки продемонстрировали преднамеренное использование ИИ во вредоносных целях. Злоумышленники постоянно меняют и совершенствуют стратегии и способы реализации атак, уделяя особое внимание применению методов ИИ, называемых кибератаками на основе ИИ, которые можно использовать в сочетании с обычными методами атак для нанесения большего ущерба. В докладе показывается современное состояние использования ИИ в кибербезопасности в виде сопоставления методов и технологий нападения и защиты. Проводится анализ данных, используемых для решения задач кибербезопасности, обобщаются существующие области применения ИИ для кибербезопасности и определяются ключевые ограничения. Рассматриваются авторские исследования в области интеллектуального мониторинга кибербезопасности и обнаружения кибератак.

**Ключевые слова:** кибербезопасность, защита информации, кибератака.

---

\* Исследования выполняются при финансовой поддержке гранта РФФ № 21-71-20078.

УДК 004.8

## **ИНТЕГРАЦИЯ ИНЖЕНЕРИИ ЗНАНИЙ И КОНСТРУИРОВАНИЯ ПРИЗНАКОВ МАШИННОГО ОБУЧЕНИЯ ПРИ ПРОЕКТИРОВАНИИ ПРИКЛАДНЫХ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

Н.Г. Ярушкина (*jng@ulstu.ru*)

Ульяновский государственный технический университет,  
Ульяновск

Доклад посвящен эффективной интеграции методов инженерии знаний и методов конструирования признаков машинного обучения, оценки наборов данных, формирования предметных рекомендательных систем, предиктивной аналитики с учетом требований доверительного искусственного интеллекта. Формирование множества обучающих данных имеет принципиально важное значение для успешного применения машинного обучения в решении задач искусственного интеллекта. В настоящее время развиваются методы машинного обучения с участием человека, которые позволяют использовать научный потенциал инженерии знаний для построения моделей и конструирования признаков. В докладе продемонстрирована эффективность предлагаемого подхода на примере ряда разработанных прикладных систем искусственного интеллекта.

**Ключевые слова:** инженерия знаний, машинное обучение, прикладные системы искусственного интеллекта.

---

СЕКЦИЯ 1

**ИНЖЕНЕРИЯ ЗНАНИЙ,  
КОГНИТИВНЫЕ  
ИССЛЕДОВАНИЯ  
И ИНТЕЛЛЕКТУАЛЬНЫЙ  
АНАЛИЗ ТЕКСТОВ**

**УДК 005.94**

**ПРОБЛЕМА КЛАСТЕРИЗОВАННОСТИ  
ОНТОЛОГИЧЕСКОЙ ИНФРАСТРУКТУРЫ  
ДЛЯ ОПИСАНИЯ ЭМПИРИЧЕСКИХ ДАННЫХ  
И ПРЕДЛОЖЕНИЕ ПО ЕЁ РЕШЕНИЮ\***

А.М. Беглер (*alena.begler@gmail.com*)

Высшая школа менеджмента СПбГУ, Санкт-Петербург

К настоящему моменту существует устоявшаяся онтологическая инфраструктура для описания данных эмпирических исследований. Однако её анализ показал, что она распадается на два относительно независимых кластера, что препятствует интеграции данных между кластерами. В работе предлагается подход, позволяющий осуществлять описание данных вне зависимости от того, в рамках какого кластера они были описаны. Подход основан на двух принципах: (1) описание информационных сущностей осуществляется в рамках отдельного кластера, независимого от кластеров, образованных предметными онтологиями, в том числе не использующего соответствующие онтологии верхнего уровня; (2) обеспечение горизон-

---

\* Работа выполнена при финансовой поддержке РФФИ (проект №20-07-00854).

тальной связности между кластерами и вертикальной – внутри них. Предложенный подход реализован в переработанной онтологии *empirion* и показана его применимость для связи существующих кластеров.

**Ключевые слова:** онтология, связность онтологий, онтологическая интеграция данных, онтологическая инфраструктура.

## Введение

К настоящему моменту существует огромное количество онтологических моделей предметных областей, часть из них широко используется для описания массивов данных – например, семейство онтологий *Open Biological and Biomedical Ontology (OBO) Foundry*<sup>1</sup> [Smith et al., 2007]. Возможности интеграции данных, полученных из разных источников (например, разными исследовательскими группами) в таком подходе обеспечиваются связностью онтологий, входящих в семейство: предметные онтологии в нём расширяют онтологию верхнего уровня *Basic Formal Ontology (BFO)*<sup>2</sup>, уточняя её для конкретных областей.

Такое состояние дел, однако, приводит и к проблеме: показано [Begler et al., 2022], что онтологические модели для описания эмпирических данных образуют два независимых кластера. Один кластер группируется вокруг упомянутой онтологии верхнего уровня BFO, второй – вокруг другой онтологии верхнего уровня *Semanticscience Integrated Ontology (SIO)*<sup>3</sup>. Несмотря на то, что онтологии на первый взгляд похожи – обе содержат на верхнем уровне классы для «объектов» (сущностей, представленных полностью в один момент времени, *bfo:continuant* и *sio:object*) и «процессов» (сущностей, развёртывающихся во времени, *bfo:occurrent* и *sio:process*), – их семантика различна. Так, например, одно из важных отличий SIO от BFO, на которое обращают внимание сами авторы [Dumontier et al., 2014], – определение процесса. SIO позволяет процессам иметь атрибуты (третий класс верхнего уровня в этой онтологии – *sio:attribute*), тогда как BFO не позволяет, потому что для BFO процесс это принципиально изменчивая сущность, у которой не может быть атрибутов, потому что изменение характеристик процесса и есть сущностно этот процесс. Кроме этого, SIO включает в себя лингвистическую и символическую репрезентации объектов, тогда как BFO не делает этого на-

---

<sup>1</sup> Семейство взаимосвязанных онтологий для биологических наук, <https://obofoundry.org/>.

<sup>2</sup> Домашняя страница онтологии BFO, <https://basic-formal-ontology.org/>.

<sup>3</sup> Репозиторий онтологии SIO, <https://github.com/MaastrichtU-IDS/semanticscience>.



прямую, предполагая использование расширений, так как Information Artifact Ontology (IAO)<sup>4</sup> для информационных сущностей и Simple Knowledge Organization System (SKOS)<sup>5</sup> для лексических. Поэтому интеграция данных, описанных в рамках разных кластеров, сопряжена со сложностями.

Вариантом решения этой проблемы может стать предлагаемый подход к разработке онтологий, разделяющий описание «предметной» составляющей данных и их фактического представления в массиве. Шаги в этом направлении уже сделаны такими онтологическими моделями как Data Catalog Vocabulary (DCAT)<sup>6</sup>, DDI-RDF Discovery Vocabulary (discovery)<sup>7</sup> и IAO. Однако проблема состоит в том, что две из трёх перечисленных онтологий – discovery и IAO – существуют в рамках выделенных кластеров, а кроме того не позволяют описать значения переменных, что является недостающим звеном для интеграции исследовательских данных [Dumontier, 2019]. В данной работе сделана попытка, во-первых, показать возможность сквозной интеграции данных, описанных в рамках существующих кластеров и обогатить моделирование переменных; во-вторых, представить целостный взгляд на онтологическую инфраструктуру, обеспечивающую возможность такого моделирования, что является важным аспектом моделирования исследовательских областей [Боргест, 2022]. Предлагаемый подход может быть полезен при проектировании прикладных онтологий, в особенности направленных на решение задач интеграции данных.

В первом разделе рассмотрена проблема связности онтологических моделей и выделено два типа связности между моделями: вертикальная и горизонтальная. Во втором разделе описана дальнейшая типизация горизонтальных связей в разрезе слоёв для описания данных (непосредственно моделируемых объектов, лексического и информационного) и в разрезе информации, необходимой для интерпретации данных: контейнер, контекст и контент. Также показан вариант реализации такого подхода в разработанной ранее онтологической модели *empirion*. В третьем разделе проанализирована применимость данного подхода в рамках выявленных кластеров онтологической инфраструктуры и показана возможность интеграции данных, описанных с использованием моделей из разных кластеров.

---

<sup>4</sup> Онтология информационных сущностей, является расширением онтологии BFO, <https://github.com/information-artifact-ontology/IAO>.

<sup>5</sup> Онтологическая модель для создания тезаурусов, <https://www.w3.org/2009/08/skos-reference/skos.html>.

<sup>6</sup> Онтология для описания каталогов данных, <https://www.w3.org/TR/vocab-dcat-2/>.

<sup>7</sup> Онтология для описания информации о массивов данных, полученных в исследовании, преимущественно опросах, <https://rdf-vocabulary.ddialliance.org/discovery.html>.

## 1. Аспекты связности онтологий в контексте интеграции массивов данных

Связность онтологий в процессе разработки обеспечивается повторным использованием других онтологий.

### 1.1. Возможности для обеспечения связности онтологий

Значительная часть литературы в области онтологического инжиниринга по повторному использованию онтологий посвящена типам повторного использования онтологий и технической реализации этого использования. Так, повторное использование онтологий может осуществляться разными способами: интеграция (integration); слияние (merging, fusion) и выравнивание (alignment); простое повторное использование без какого-либо изменения исходной онтологии [Pinto et al., 1999; Гаврилова и др., 2020; Carriero et al., 2020]. При этом под интеграцией понимается использование наборов сущностей из уже существующей онтологии в разрабатываемой, под слиянием – объединение онтологий, а под выравниванием – обеспечение ссылок между онтологиями без непосредственного повторного использования сущностей из одной онтологии в другой. Можно сказать, что когда мы говорим о расширении онтологий верхнего уровня предметными, то речь идёт об интеграции или слиянии, а когда об обеспечении связей между разными предметными онтологиями – о выравнивании.

Данная тройственность отражена и в возможностях технической реализации повторного использования онтологий [Carriero et al., 2020]: прямое использование (direct reuse) всей онтологии с помощью *owl:imports*; прямое использование отдельных сущностей онтологии с помощью ссылки *rdfs:isDefinedBy*; не прямое повторное использование (indirect reuse) через выравнивание онтологий, например, с помощью *rdfs:subClassOf* или *owl:equivalentProperty*.

### 1.2. Вертикальная и горизонтальная связность

Другой аспект повторного использования отражает «смысловой» уровень связности онтологий. Принятое деление онтологий на онтологии верхнего уровня, предметных областей и приложений предполагает использование онтологий более высокого уровня при разработке онтологий более низкого уровня (например, [Suárez-Figueroa et al., 2012]) – это обеспечивает **вертикальную связность**.

Кроме этого возможна **горизонтальная связность**: использование онтологий одних предметных областей при разработке онтологий других предметных областей. Например, многим предметным областям необходима информация о единицах измерения. Однако в отличие от расширения

онтологии верхнего уровня здесь связи между онтологиями сложнее. На интуитивном уровне понятно, что есть разница между расширением, например, онтологии биологических сущностей (1) словарём разных названий содержащихся в этой онтологии сущностей; (2) онтологиями единиц измерения; (3) онтологиями, содержащими информацию об авторстве и версиях разрабатываемой онтологии. Во всех трёх случаях речь идёт не о расширении онтологии более высокого, по сравнению с разрабатываемой, уровня абстракции, а о связности онтологий похожего уровня абстракции:

- В первом случае речь идёт об использовании лексической модели параллельно с моделью явлений реального мира. Здесь речь идёт о тройственности знака, смысла и значения, отражённой в семантическом треугольнике Огдена и Ричардса [Ogden et al., 1923] и обеспечиваемой обычно моделями тезаурусов типа SKOS.
- Второй пример – использование частей одной предметной онтологии при разработке другой.
- Третий – применение модели метаданных.

В случае интеграции данных из разных массивов важно понимание и того, в каких единицах измерения представлены переменные, и того, кем и когда получен массив. Данный разрез рассмотрен в литературе о метаданных, освещённой в следующем разделе.

## **2. Подход к интеграции данных, независимый от предметных областей**

Литература, посвящённая моделям метаданных предлагает разделение блоков метаданных по их целевому назначению. Представляется, что этот подход можно использовать для основанной на онтологиях интеграции данных.

### **2.1. Схемы метаданных как инструмент для интеграции данных**

В литературе о метаданных зачастую разделяют описательные, структурные и административные элементы [Greenberg, 2005; Mendez & van Hooland, 2013; Riley, 2017]. Разные авторы предлагают несколько отличающееся разделение типов элементов, но, обобщая перечисленные работы, можно понимать под описательными элементами те, что обеспечивают понимание данных (название, кем и когда получены); под административными – обеспечивающие их использование (лицензия, идентификация), под структурными – состав и структуру данных (формат и кодирование данных, переменные и их связи). Все эти элементы необходимы для описания данных и, значит, должны быть включены в схему метаданных. Отметим здесь, что описательные и административные элементы метаданных относятся к массиву данных в целом, тогда как структурные час-

точно – к массиву данных в целом, а частично к его составным частям. Также, именно структурные элементы обеспечивают пользователям понимание переменных в массиве.

Опросы исследователей, непосредственно использующих массивы данных в своей работе [Faniel et al., 2019; Koesten et al., 2020], выявили следующие элементы, необходимые пользователям для понимания данных: контекст получения данных (процедура исследования, происхождение данных), особенности массива данных (общие свойства массива данных, переменные и их связи), контекст массива данных в предметной области (ссылки на стандартные описания, другие исследования в области, возможный анализ и интерпретация данных).

## 2.2. Семейство онтологий для интеграции данных

Развивая рассмотренное направление мысли можно предложить такой набор типов сущностей, составляющий онтологическую модель для описания данных, который позволит обеспечить интеграцию данных:

- Описывающие **контекст данных**: процедуру получения данных, элементы схем метаданных общего назначения для сохранения информации об авторстве, лицензировании и так далее.
- Описывающие **контейнер данных**, т.е., непосредственно само физическое представление массива, в котором содержатся данные.
- Описывающие **контент данных**: содержательная часть, обеспечивающая понимание и интерпретацию значений переменных.

Такой взгляд на онтологическую модель для описания данных существует одновременно с описанием непосредственно предметной области, и представляет собой дополнительный разрез, описывающий информационные сущности – аналогично тому, как параллельно с моделью предметной области существует лексическая модель этой же области. Точно так же, как модели предметных областей расширяют онтологии верхнего уровня (например, Gene Ontology, расширяющая, как и другие онтологии семейства ОВО Foundry, онтологию VFO), модели для описания данных могут образовывать вертикальные связи и формировать отдельный кластер (так, онтология discovery для интеграции данных опросов является расширением модели DCAT).

В описанном подходе онтологии для интеграции данных должны сохранять с имеющимися инфраструктурными кластерами горизонтальную связь, но строиться независимо от моделей предметных областей, избегая с ними вертикальных связей. Пример реализации такого подхода представлен на рис. 1.

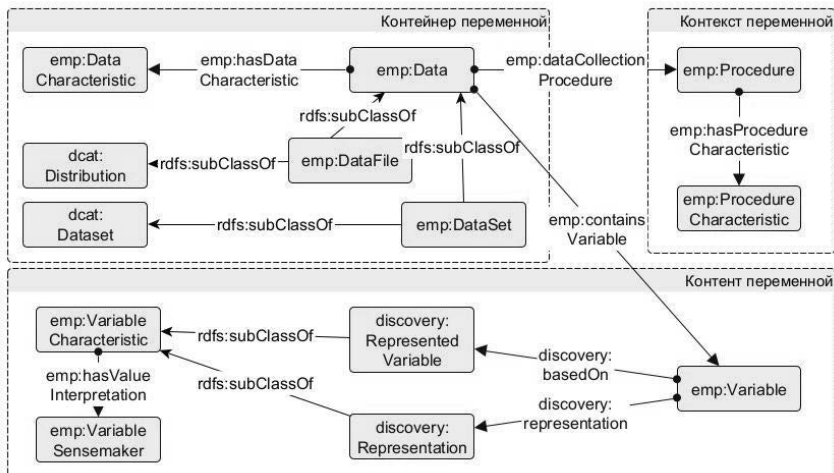


Рис. 1. Реализация подхода контейнер-контент-контекст в онтологии *empirion* как части кластера для описания данных

На данном рисунке созданная ранее онтология для описания переменных в массивах данных эмпирических исследований *empirion*<sup>8</sup> [Беглер, Лещева, 2021] переработана как расширение онтологии DCAT, повторно использует сущности онтологии *discovery* и не содержит никакие предметные классы, но предоставляет возможность для расширения такими классами. Такой подход отличается от уже представленных моделей в четырёх аспектах:

1. Онтология содержит вертикальные связи с онтологией для описания данных DCAT и горизонтальные связи с онтологией для описания переменных *discovery* образовывая отдельный кластер инфраструктуры, независимый от кластеров предметных областей и таким образом позволяющий интегрировать данные, описанные в рамках любой предметной онтологии.
2. Онтология предоставляет возможность описать одновременно и физическое представление данных (Контейнер переменной на рисунке 1), и процесс их получения (Контекст переменной), и информацию, необходимую для их интерпретации (Контент переменной).

<sup>8</sup> Онтология переменных для интеграции данных эмпирических исследований, <https://github.com/jimijimiyo/empirion/>.

3. Любой из классов онтологии *empirion* может быть расширен классами предметной области, при этом сохраняются все связи, необходимые для интерпретации значений переменных.
4. Данный подход не вступает в конфликт с лексической моделью и моделями метаданных, позволяя сохранять информацию о переменных в дополнение к словарю предметной области и другим метаданным – примеры будут показаны в следующем разделе.

### 3. Анализ применимости подхода

Предлагаемый подход был проверен на выявленных ранее [Begler et al., 2022] кластерах инфраструктуры для описания исследовательских данных.

Первый кластер образуется вокруг онтологии BFO и сфокусирован в первую очередь на представлении сущностей предметных областей, для описания информационного разреза в данном кластере используется онтология IAO, расширяющая BFO (рис. 2).

Видно, что в данном кластере никак не отражена информация о переменных, кроме того, нет никаких ссылок на специализированные модели представления данных. В таком случае установление связей с *emp:DataSet* обеспечивает одновременно и связь с кластером для описания данных через DCAT (так как *emp:DataSet* является подклассом *dcat:dataset*), а также с переменными в массивах, так как *emp:DataSet rdfs:subClassOf emp:Data* и *emp:Data emp:containsVariable emp:Variable* и далее с информацией, необходимой для интерпретации переменной (см. рис. 1). Если одновременно рассмотреть некоторые классы онтологии BFO как подклассы процедуры получения данных *emp:Procedure*, то сохранится и связь с контекстом.

Второй кластер группируется вокруг онтологии SIO и уже содержит связь с моделью данных DCAT. В этом кластере также уже существует информация о представлении переменных и необходимо только дополнить её информацией, необходимой для интерпретации значений переменных (рис. 3).

В обоих кластерах сущности онтологии *empirion* (1) позволяют интегрировать массивы данных в случаях, когда они описаны в рамках разных инфраструктурных кластеров за счёт горизонтальных связей; (2) дополняют кластер для описания информационных сущностей, обеспечивая сохранение информации, необходимой для интерпретации значений переменных за счёт вертикальных связей.

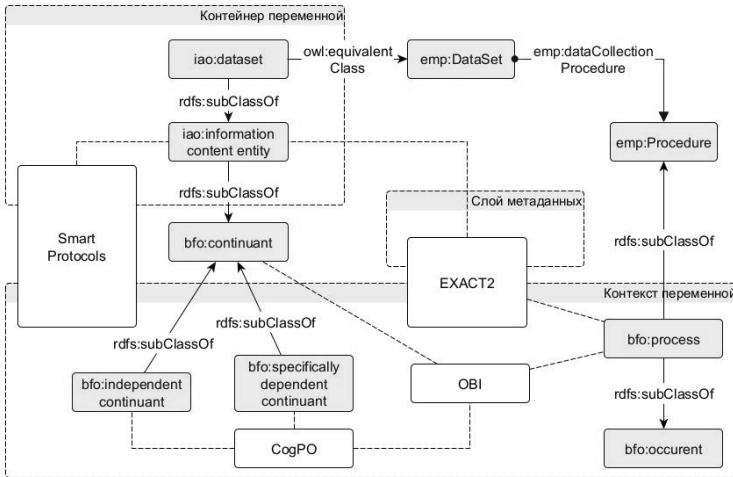


Рис. 2. Часть кластера, образованного вокруг онтологии BFO (показана только часть сущностей). В белых прямоугольниках указаны онтологии, расширяющие соответствующие классы

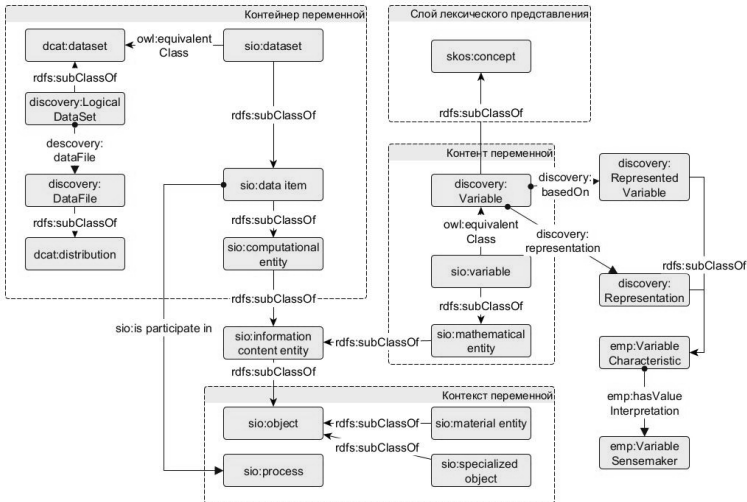


Рис. 3. Часть кластера, образованного вокруг онтологии SIO (показана только часть сущностей)

## Заключение

Данная работа появилась в результате попыток обеспечить интеграцию данных, полученных в эмпирических исследованиях разными исследовательскими группами. Первая проблема – обеспечение интерпретации значений переменных – была решена при первичном построении онтологии *empirion*. Однако, в процессе изучения экосистемы онтологических моделей для описания данных была выявлена новая проблема – наличие в этой экосистеме как минимум двух независимых кластеров. Такое состояние дел делает онтологию для интеграции данных, разработанную в рамках одного из кластеров, не в полной мере пригодной для интеграции данных.

В поисках решения этой новой проблемы был сформулирован описанный подход, в котором предлагается относительно независимое построение моделей предметных областей и информационных сущностей, по аналогии с тем как это всё чаще реализуется, например, для лексических моделей. Важный момент здесь – разделение вертикальной и горизонтальной связности между онтологиями.

Вертикальная связность предполагает разработку онтологий более низкого уровня абстракции как расширения онтологий верхнего уровня. Так, для кластера, описывающего информационные сущности, в который входит онтология *empirion*, такой онтологией служит DCAT. Горизонтальная связность существует между кластерами, и в рассматриваемом случае позволяет гармонизировать переменные, описанные в рамках разных кластеров и, таким образом, способствовать интеграции массивов данных.

## Список литературы

- [Begler et al., 2022] Begler A., Anufriev G., Leshcheva I. Ontology of Experimental Variables as an Extension of Infrastructure for Behavioral Research Data FAIRification // *Metadata and Semantic Research*. Springer International Publishing. – 2022. – P. 268-279.
- [Carriero et al., 2020] Carriero V.A. et al. The Landscape of Ontology Reuse Approaches. – 2020. – № section 1. – P. 1-18.
- [Dumontier et al., 2014] Dumontier M. et al. The SemanticScience Integrated Ontology (SIO) for biomedical research and knowledge discovery // *J. Biomed. Semantics*. – 2014. – Vol. 5, No. 1. – P. 1-11.
- [Dumontier, 2019] Dumontier M. Accelerating biomedical discovery with an Internet of FAIR data and services. – 2019.
- [Greenberg, 2005] Greenberg J. Understanding Metadata and Metadata Schemes // *Cataloging & Classification Quarterly*. – 2005. – Vol. 40, No. 3-4. – P. 17-36.
- [Faniel et al., 2019] Faniel I.M., Frank R.D., Yakel E. Context from the data reuser's point of view // *Journal of Documentation*. – 2019. – Vol. 75, No. 6. – P. 1274-1297.



- [**Koesten et al., 2020**] Koesten L. et al. Dataset Reuse: Toward Translating Principles to Practice // *Patterns Prejudice*. Elsevier Inc. – 2020. – Vol. 1, No. 8. – P. 100136.
- [**Mendez & van Hooland, 2013**] Méndez E., van Hooland S. Metadata Typology and Metadata Uses // *Handbook of Metadata, Semantics and Ontologies*. – WORLD SCIENTIFIC, 2014. – P. 9-39.
- [**Ogden et al., 1923**] Ogden C.K., Richards I.A. The Meaning of Meaning. – A Harvest Book, 1923.
- [**Pinto et al., 1999**] Pinto S., Gómez-Pérez A., Martins J.P. Some Issues on Ontology Integration // *Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5)*. – CEUR Workshop Proceedings, 1999.
- [**Riley, 2017**] Riley J. Understanding Metadata: What Is Metadata, and What is it for? // *NISO Primer*. – 2017. – P. 42.
- [**Smith et al, 2007**] Smith B. et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration // *Nat. Biotechnol.* – 2007. – Vol. 25, No. 11. – P. 1251-1255.
- [**Suárez-Figueroa et al., 2012**] Suárez-Figueroa M.C., Gómez-Pérez A., Fernández-López M. The NeOn Methodology for Ontology Engineering // *Ontology Engineering in a Networked World* / ed. by Suárez-Figueroa M.C. et al. – Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. – P. 9-34.
- [**Беглер, Лещева, 2021**] Беглер А.М., Лещева И.А. Онтология empirion для описания переменных в массивах данных, полученных в экспериментальных исследованиях // *Девятнадцатая Национальная конференция по искусственному интеллекту с международным участием КИИ-2021 (11–16 октября 2021 г.): Труды конференции / под ред. В.В. Борисова, Б.А. Кобринского*. – Ростов-на-Дону; Таганрог: Изд-во ЮФУ, 2021. – С. 87-98.
- [**Боргест, 2022**] Боргест Н.М. Онтология проектирования научного направления: формирование, развитие, примеры // *Онтология проектирования*. – 2022. – Т. 12, № 2 (44). – С. 136-157.
- [**Гаврилова и др., 2020**] Гаврилова Т.А., Кудрявцев Д.В., Муромцев Д.И. Инженерия знаний. Модели и методы. – СПб.: Лань, 2016. – 324 с.

УДК 004.822

## ОНТОЛОГИЧЕСКИЙ ИНЖИНИРИНГ В СЛОЖНЫХ СИСТЕМАХ НА ОСНОВЕ МЕТА-АССОЦИАТИВНЫХ ГРАФОВ\*

В.В. Борисов (*vbor67@mail.ru*)  
филиал ФГБОУ ВО «Национальный исследовательский университет «МЭИ» в г. Смоленске, Россия

А.Е. Мисник (*anton@misnik.by*)  
МОУ ВО «Белорусско-Российский университет»,  
Республика Беларусь, Могилев

В работе предложена разновидность метаграфов, ориентированная на применение в онтологическом инжиниринге сложных систем, характеризующихся сложностью структуры, многокомпонентностью, наличием функциональных подсистем, решающих различные целевые задачи, большим количеством параметров, характеризующих процессы системы. Онтологический инжиниринг ставит своей целью обеспечение принятия качественных управленческих решений за счет повышения уровня интеграции необходимой информации, улучшения возможности поиска по базам данных и знаний, обеспечения возможности совместной работы с базой знаний, что обеспечивается единым семантическим описанием прикладной онтологии. Мета-ассоциативные графы позволяют обеспечить гибкое представление онтологии предметной области сложной системы, пригодное для моделирования системных процессов и обеспечения функционирования информационно-аналитических процессов в системе, минимизируя, в ситуациях где нет необходимости в детализации, необходимое количество шагов для получения необходимой информации, но, при наличии потребности в детализированных данных, обеспечивая возможности для получения всех связанных данных. Включение процессной составляющей в онтологическую модель системы позволяет повысить общую гибкость моделирования процессов.

**Ключевые слова:** онтологический инжиниринг, метаграфы, мета-ассоциативные графы.

---

\* Исследование выполнено за счет гранта Российского научного фонда № 22-61-00096, <https://rscf.ru/project/22-61-00096/>.

## Введение

Эффективность управления сложными системами во многом зависит от правильно спроектированных и отлаженных системных и, сопровождающих их, информационно-аналитических процессов [Borisov et al, 2020a], [Bobryakov et al, 2020].

Онтологический подход обладает рядом ключевых преимуществ при проектировании и сопровождении процессов в сложных системах, поскольку особенностью таких систем является высокий уровень интеграции процессов и совместное использование данных и знаний [Meyer, 1988], [Mrochek et al, 2020].

Особенностями сложных систем являются:

- сложность структуры, многокомпонентность, наличие функциональных подсистем, решающих различные целевые задачи, большое количество параметров, характеризующих процессы системы;
- динамическое изменение структуры и параметров системы;
- неполная информация о функционировании и состоянии системы;
- разнообразие внешних и антропогенных воздействий на систему;
- наличие сложных нелинейных связей между параметрами;
- сложность принятия управленческих решений по повышению эффективности системы;
- ограниченные возможности для экспериментальных исследований системы и происходящих процессов;
- невозможность использования единого подхода к созданию моделей производственно-технических и информационно-аналитических процессов, происходящих в таких системах [Leveson et al, 1994], [Горбунов-Посадов, 1999].

Проведенные исследования позволили обосновать применение мета-ассоциативных графов для онтологического инжиниринга сложных систем. Определены требования к инструментарию, интегрирующему онтологическое и метаграфовое представления в рамках единой программно-инструментальной среды, что не достигается традиционным объединением онтологий.

### 1. Особенности онтологий сложных систем

Онтология описывает понятия в предметной области, а также отношения, которые существуют между этими понятиями. Таким образом, базовыми характеристиками онтологии являются иерархия понятий/объектов, которая устанавливается с помощью различных семантических связей, названия которых выбраны исходя из принятой терминологии предметной области, с целью преодоления семантического разрыва между экспертом и инженером по знаниям [Wald'en et al, 1995], [Rak, 2020].

Онтологический инжиниринг подразумевает:

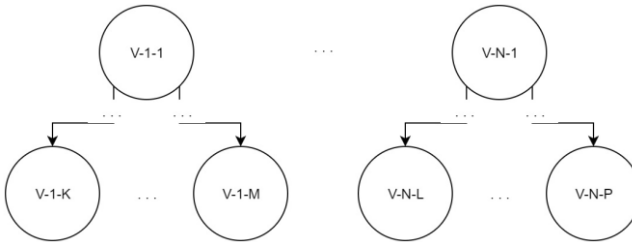
- определение классов понятий в онтологии;
- руководство по таксономии классов;
- разработка структур понятий и ситуаций;
- определение свойств понятий и значений этих свойств;
- процедуры вывода и преобразования ситуаций [Kendall et al, 2020].

Для практического использования на завершающих этапах разработки прикладная онтология должна включать в себя так называемые экземпляры (сущности) созданных классов с конкретными значениями их свойств, описывающих реальное отображение рассматриваемой прикладной онтологии [Aguado De Sea et al, 2008], .

Сложность структуры, многокомпонентность и наличие функциональных подсистем внутри сложной системы, естественным образом влияют на представление её онтологии. В общем случае онтологию сложной системы можно рассматривать как спецификацию процессов ее построения и эксплуатации, вне зависимости от методов ее реализации и также вне зависимости от применяемых для её реализации технологий [Borisov et al, 2020b].

## 2. Графы, как средство представления онтологии сложной системы

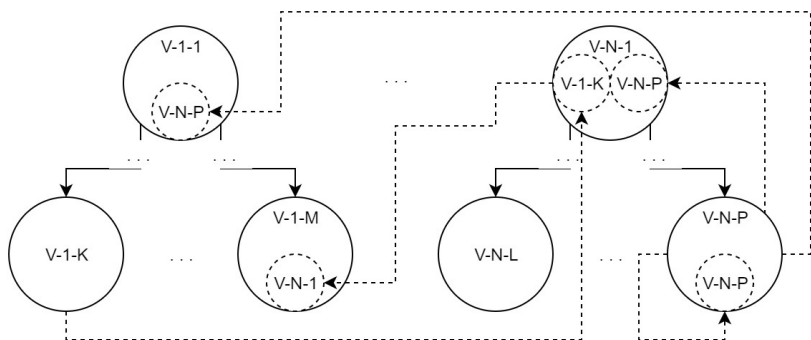
При попытке представить онтологию сложной системы в виде классических графов, может показаться, что она состоит из абсолютно независимых сегментов, каждый из которых является классическим графом (рис. 1).



*Рис. 1. Попытка представления онтологии сложной системы в виде классических графов*

Однако, наличие сложных нелинейных связей между параметрами и сущностями сложной системы не позволяют использовать классические графовые структуры [Jensen et al, 1994]. Онтология сложной системы может

быть достаточно сложно организована и сущности одного, на первый взгляд независимого графа, могут включаться в сущности этого же или другого графа (рис. 2).



*Рис. 2. Иллюстрация зависимостей между сущностями онтологии сложной системы*

Очевидно, что инструментарий классических графов не позволяет формализовать подобные взаимосвязи между сущностями онтологии сложной системы [Tanenbaum et al, 2002], [Kutuzov et al, 2020].

Необходимость повышения адекватности онтологии приводит к поиску инструментария для использования не только бинарных отношений между вершинами графа, но и отношений, объединяющих более двух вершин. При этом в качестве модели может использоваться гиперграф. Ребро гиперграфа, соединяющее набор вершин, соответствует определенному отношению между данным набором вершин. Это, например, может быть совокупность компонентов системы, участвующих в выполнении определенной функции, или совокупность данных, описывающих определенный объект, свойства, события, или совокупность элементов, объединенных в одно целое [Гапанюк, 2020].

В настоящее время гиперграфы широко используются в тех областях, где требуется анализ сложных связей между вершинами графа, в частности, при проектировании аппаратных средств. Однако в силу определения гиперграф не обладает достаточно удобными возможностями для моделирования иерархий. В отличие от гиперграфа метаграф представляет собой модель иерархического объекта с разным уровнем обобщения [Самохвалов и др., 2015].

Метаграф – это набор направленных отображений множества в множество. Его многообещающими приложениями являются моделирование отношений данных, каждое из которых рассматривается как отображение набора ключевых элементов на набор данных, и моделирование задач рабочего процесса, каждый из которых можно рассматривать как отображение набора элементов, входных документов в набор выходных документов (рис. 3).

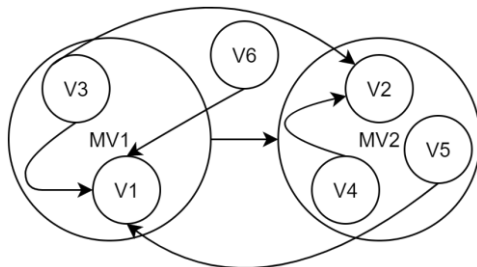


Рис. 3. Пример метаграфа

Метаграф – определяется тремя множествами:

$$S = \{V, MV, E\},$$

где  $V$  – множество вершин метаграфа;  $MV$  – множество метавершин;  $E$  – множество ребер. Метавершина метаграфа – это вершина, включающая множество вершин  $v$ ,  $v \in V$ .

Вершина метаграфа может быть атрибутивной, т.е. с ней может быть связано произвольное количество атрибутов.

$$v_i = \{A\}, v_i \in V,$$

где  $v_i$  – вершина метаграфа;  $A$  – множество атрибутов.

Ребро метаграфа характеризуется исходной и целевой вершинами и направленностью и соединяет две вершины, в отличие от гиперребра гиперграфа, которое может охватывать несколько вершин. Ребро может соединять как вершины, так и метавершины метаграфа [Basu et al, 2007].

### 3. Мета-ассоциативные графы

При описании онтологии сложной системы, метаграф с атрибутивными вершинами, в ряде случаев, всё ещё не является идеальным инструментом. Требования к динамичности изменения структуры и параметров системы, с точки зрения онтологического моделирования процессов в

сложной системе, делают отличия между атрибутами и вершинами внутри метавершины довольно зыбкими. Так и вершина метаграфа, фактически, является вырожденным случаем рассмотрения мета-вершины без включенных в неё вершин. Кроме того, у классического метаграфа отсутствуют естественные механизмы идентификации событий и реакции на них [Мисник, 2022].

Предлагается обобщить понятия вершины и метавершины до понятия узла метаграфа ( $N$ ) и дополнить определение узла именем, наборами событий и методов, принадлежащих узлу:

$$N = \{I, AS, EV, M\},$$

где  $I$  – имя узла;  $EV$  – набор событий, связанных с узлом метаграфа;  $M$  – набор методов, связанных с узлом метаграфа.  $AS$ , трансформируется из множества атрибутов  $A$  в множество ассоциативных атрибутов, являющихся либо обычными атрибутами, либо ссылками на узлы метаграфа:

$$AS = \{A, N^*\}.$$

Такая разновидность метаграфа позволяет повысить гибкость моделирования онтологии, в особенности, онтологий процессов сложной системы, поскольку ссылки на узлы метаграфа, фактически, мало чем отличаются от ребер метаграфа и позволяют обеспечить ассоциативные связи между узлами. Любая вершина потенциально становится метавершиной, поскольку ссылка на другой узел метаграфа предоставляет всю полноту возможностей для взаимодействия с ним. С другой стороны, при отсутствии необходимости установки ассоциативной связи, ссылка может быть представлена в виде обычного атрибута со значением равным имени узла (рис. 4).

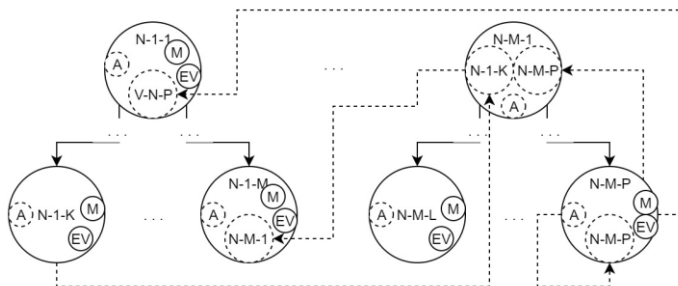


Рис. 4. Мета-ассоциативный граф

Кроме того, такой подход к построению метаграфа позволяет включать процессную составляющую в онтологическую модель системы как неотъемлемую ее часть, что позволяет гибко и экономично формировать процессные модели на основе сегментов общей онтологической модели.

#### 4. Программно-инструментальные средства онтологического инжиниринга на основе мета-ассоциативных графов

Разработана специализированная программно-инструментальная среда, в рамках которой возможно проектирование предметной онтологии системы, её системных процессов, и реализация информационно-аналитических процессов [Misnik et al, 2019].

Рассмотрим сравнительно простую задачу организации судейства на спортивных соревнованиях. Для простоты, ограничимся рассмотрением процессов, ориентированных только на судей и главных судей соревнований.

Иерархическая структура онтологии представлена на рис. 5.

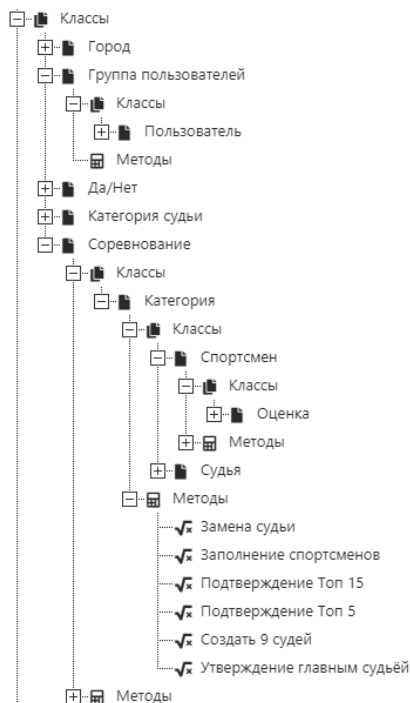


Рис. 5. Онтология предметной области «Соревнования»



Классы (узлы мета-ассоциативного графа) «Город», «Да/Нет», «Категория судьи», являются независимыми графами внутри онтологии предметной области и, по сути, выполняют функции справочников.

Классы «Город» и «Категория судьи» связаны через ассоциативные связи с классом «Пользователь» (рис. 6).

<input type="button" value="+ Добавить"/> <input type="button" value="✎ Редактировать"/> <input type="button" value="✕ Удалить"/> <input type="button" value="↕"/>							
Name	uid	Data type	Array	Unique	Require	Method	LinkType
E-mail	00000000-0000-0000-0000-000000000004	text		Да	Да		
Фамилия	87f189fe-0253-4f43-af82-f1f6728d8606	text		Да			
Имя	01fe567a-fdb0-43f8-b20a-17245559f7ad	text		Да			
Отчество	b08c28e8-9305-49d2-a421-49bb9c49c3be	text		Да			
Город	6eede639-c135-4213-a298-0899dc3b7ac8	object					Да
Категория судьи	75055304-9fc7-4c05-a9a1-2b7b90abc529	object					Да

Рис. 6. Класс «Пользователь» и его ассоциативные связи

Класс «Оценка», представляет интерес с той точки зрения, что не содержит ни одного независимого атрибута, а только ассоциативные связи с другими классами (рис. 7)

<input type="button" value="+ Добавить"/> <input type="button" value="✎ Редактировать"/> <input type="button" value="✕ Удалить"/> <input type="button" value="↕"/>							
Name	uid	Data type	Array	Unique	Require	Method	LinkType
Судья	c7b20ea0-9f14-40d5-8f11-e4d1a9e67363	object		Да			Да
Топ 5	b6b76ff5-05fa-417b-bfff0-a7e63be930b8	object					Да
Топ 15	8ea9b6da-16c7-452a-a8a0-f93d159f6788	object					Да
Утверждена Топ 15	c2050348-7e77-47d6-9fe7-39bf5d6eea78	object					Да
Номер судьи	d61d6f36-0a4b-4200-ac49-6d25f6462fe7	object					Да
Утверждена Топ 5	cefadb82-a9bc-4c73-9d91-049e030d1e21	object					Да

Рис. 7. Класс «Оценка»

На рис. 5 видно, что класс «Соревнование» включает ряд методов, необходимых для функционирования процессов системы.

Методы реализуются средствами разработанного внутрисистемного языка программирования, предназначенного для непосредственного использования онтологии и данных системы.

Методы могут быть инициированы при возникновении системных событий, либо вызваны вручную. На рис. 8 и 9 приведен общий вид редактора внутреннего языка программирования и диалоговое окно настройки

методов. События могут срабатывать как синхронно, так и в асинхронном режиме, в порядке очереди. Модуль редактора обладает функционалом отслеживания изменений и версииности, а также включает отладчик.

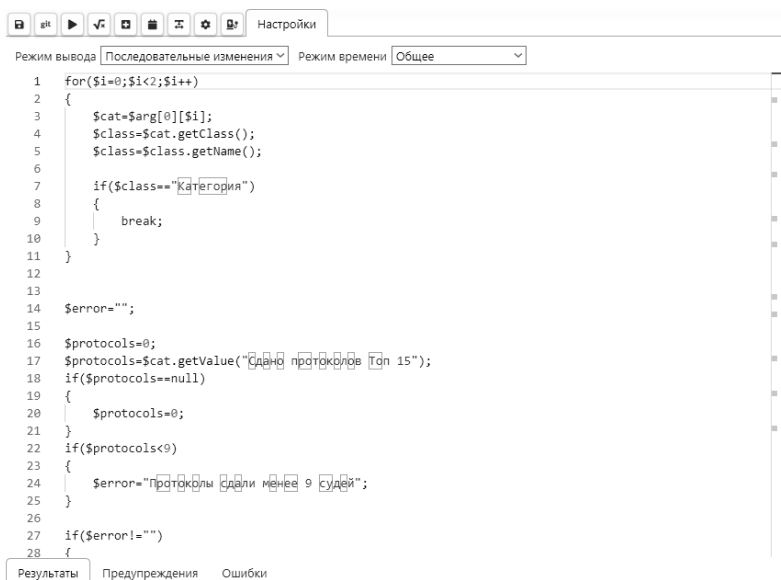


Рис. 8. Редактор внутреннего языка программирования

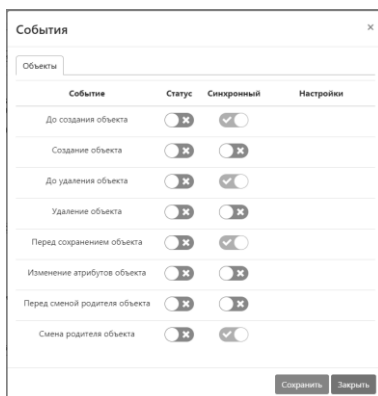
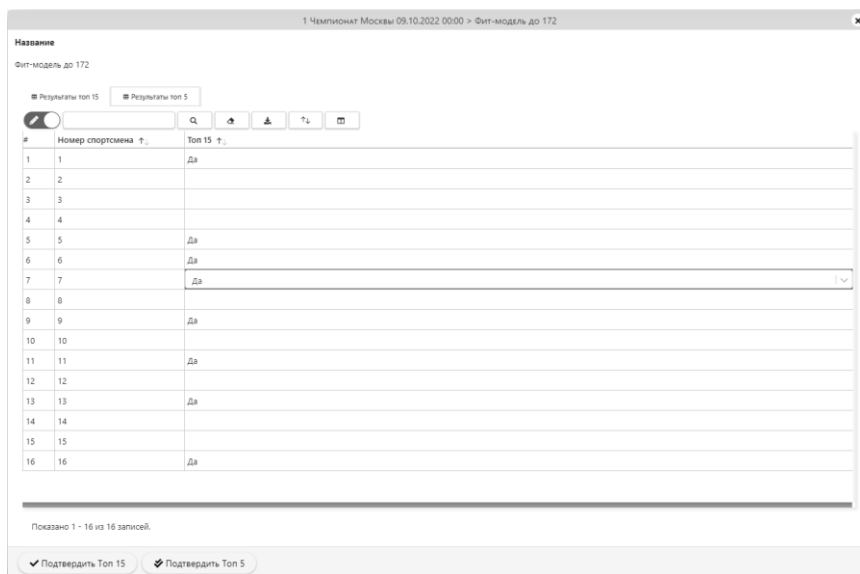


Рис. 9. События, иницирующие методы

Рассмотрим типовые процессы Судьи и Главного судьи, реализованные в единой онтологии предметной области.

Задачей Судьи является выбор спортсменов в категории, которую он судит (рис. 10). В данном случае, ассоциативная связь от класса «Оценка» к классу «Пользователь» используется для идентификации принадлежности оценки конкретному судье, а также для разграничения доступа судей к их оценкам. Экранная форма содержит методы «Подтвердить Топ 15» и «Подтвердить Топ 5», которые Судья должен вызвать вручную, после выставления оценок.



*Рис. 10. Выставление оценок Судьёй*

Для Главного судьи система агрегирует данные по выставленным судьями оценкам (рис. 11). Программно-инструментальная среда позволяет развернуть часть данных таблицы для обеспечения более удобного представления. До поворота таблица на рис. 11 была близка по виду к таблице на рис. 10, с дополнительными атрибутами.

Таким образом, одна и та же структура данных, представляющая предметную онтологию, и использующая мета-ассоциативные связи может быть представлена по-разному, в зависимости от потребностей моделирования и сопровождения конкретного процесса.

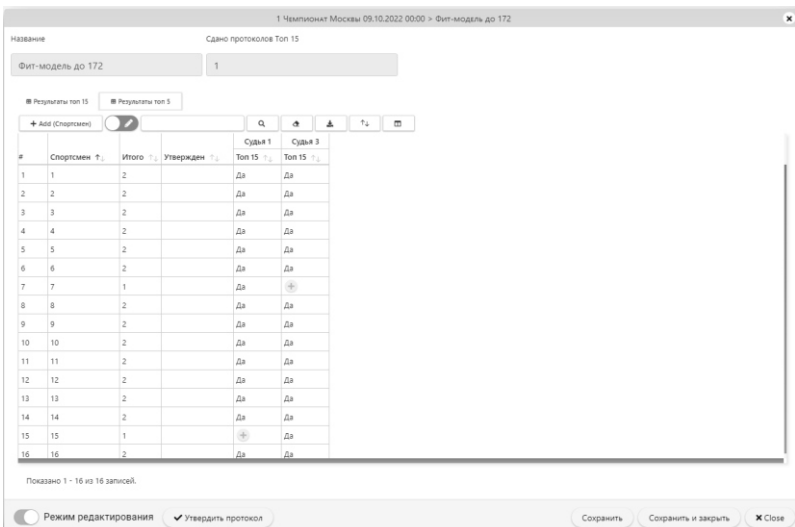


Рис. 11. Утверждение протокола Главным судьёй

## Заключение

Моделирование и выполнение процессов, направленных на поддержку управленческих решений в сложной системе, фактически, невозможно без построения онтологической модели её предметной области и процессов.

Процесс построения онтологии сложной системы является нетривиальной задачей и требует комплексного подхода – онтологического инжиниринга (для эффективного решения задачи онтологического инжиниринга требуются универсальные средства представления иерархии, атрибутов, событий и методов внутри онтологии).

Предложенная новая разновидность мета-ассоциативных графов, предназначенная для онтологического инжиниринга сложной системы решает поставленные задачи и показывает свою эффективность в реальных приложениях.

Онтологическая модель предметной области и, построенные на её базе, модели системных и информационно-аналитических процессов образуют основу для управления в сложных системах, которое с каждым годом усложняется с увеличением сложности, скорости и объёмов потоков данных, сопровождающих процессы, протекающие в них. Предложенный подход позволяет расширять и гибко перестраивать существующую онтологию в соответствии с изменяющимися требованиями.

Разработанное специализированное программно-инструментальное обеспечение для онтологического инжиниринга было апробировано и показало свою эффективность при решении реальных задач.

### Список литературы

- [**Aguado De Cea et al, 2008**] Aguado De Cea, G., Gómez-Pérez A., Montiel-Ponsoda E., and Suárez-Figueroa M.C. Natural Language-Based Approach for Helping in the Re-use of Ontology Design Patterns // In Proceedings of the 16th International Conference on Knowledge Engineering (EKAW). – 2008. – P. 32-47.
- [**Basu et al, 2007**] Basu A., Blanning R. Metagraphs and their applications. – Springer, 2007. – 174 p.
- [**Bobryakov et al, 2020**] Bobryakov A.V., Borisov V.V., Misnik A. E., Prokopenko S.A. Approaches to the Implementation of Information-Analytical Processes in Complex Technical-Organizational Systems // V International Conference on Information Technologies in Engineering Education. – 2020.
- [**Borisov et al, 2020a**] Borisov V., Fedulov A, Kruglov V. Fuzzy models and networks. – M.: Hot line – Telecom, 2018. – 284 p.
- [**Borisov et al, 2020b**] Borisov V., Lufarov V. Forecasting of Multidimensional Time Series Basing on Fuzzy Rule-Based Models // Proceedings of the 21th International Conference: “Complex Systems: Control and modeling problems”. – 2019. – Vol. 2. – P. 217-220.
- [**Jensen et al, 1994**] Jensen C., Snodgrass R., Soo M. Unifying Temporal Data Models Via a Conceptual Model // Information Systems. – 1994. – Vol. 19, No. 7. – P. 513-547.
- [**Kendall et al, 2020**] Kendall E.F., McGuinness D.L., Ding Y., Groth P. Ontology Engineering, – Morgan & Claypool, 2019.
- [**Kutuzov et al, 2020**] Kutuzov V., Misnik A., Prakupenka S. Development of Information-Analytical Processes in Cyber-Physical Systems Based on Neural-Fuzzy Petri Nets // International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon), Vladivostok, 2020.
- [**Leveson et al, 1994**] Leveson N.G., Heimdahl M., Hildreth H., Reese J.D. Requirements specification for process-control systems. Software Engineering // IEEE Transactions on. – 1994. – 20 (9). – P. 684-707.
- [**Meyer, 1988**] Meyer B. Object-oriented software construction. – Vol. 2. Prentice hall New York, 1988.
- [**Misnik et al, 2019**] Misnik A., Krutolevich S., Prakupenka S., Lukjanov E. Methodology for Development of Industrial Analytical Systems for Data Collection and Processing // Proceedings of the 14th International Conference on Interactive Systems: Problems of Human-Computer Interaction. Ulyanovsk, Russia, September 24-27, 2019. – P. 223-231.
- [**Mrochek et al, 2020**] Zakharchenkov K.V., Mrochek Zh. A., Mrochek T.V. Algorithm for solving a multicriteria problem of planning the production of pipes and fittings // System Analysis and Applied Informatics. – 2018.

- [**Rak, 2020**] Rak T. Modeling Web Client and System Behavior // MDPI/Information. – 2020.
- [**Tanenbaum et al, 2002**] Tanenbaum Andrew S., Maarten van Steen. Distributed Systems. Principles and paradigms. – Prentice Hall, Inc., 2002.
- [**Wald'en et al, 1995**] Wald'en K., Nerson J.M. Seamless object-oriented software architecture. – Prentice-Hall, 1995.
- [**Гапанюк, 2020**] Гапанюк Ю.Е., Ханмурзин Т.И., Костян А.А. Использование метаграфового подхода в концептуальном моделировании // Динамика сложных систем – XXI век. – 2020. – Т. 14, № 2. – С. 54-62. – Doi: 10.18127/j19997493-202002-06. – EDN WPCGHW.
- [**Горбунов-Посадов, 1999**] Горбунов-Посадов М.М. Расширяемые программы. – М.: Полиптех, 1999. – 336 с.
- [**Мисник, 2022**] Мисник А.Е. Применение метаграфов для онтологического инжиниринга сложных систем // Прикладная информатика. – 2022. – Т. 17, № 2 (98). – С. 120-132. – Doi: 10.37791/2687-0649-2022-17-2-120-132.
- [**Самохвалов и др, 2015**] Самохвалов Э.Н., Ревунков Г.И., Гапанюк Ю.Е. Использование метаграфов для описания семантики и прагматики информационных систем // Вестник МГТУ им. Н.Э. Баумана. Серия «Приборостроение». – 2015. – № 1 (100). – С. 83-99.

## РАСШИРЕНИЕ МЕТАГРАФОВОЙ МОДЕЛИ НА ОСНОВЕ ПРОТОГРАФОВОГО ПОДХОДА \*

Ю.Е. Гапанюк (*gapyu@bmstu.ru*)  
МГТУ им. Н.Э. Баумана, Москва

Основная проблема, которая решается в данной работе, связана с тем, что в предыдущих вариантах формализации метаграфовой модели недостаточно детально рассматривался вопрос о том, как элементы модели соединяются в единый метаграф. Рассматриваются основные положения аннотируемой метаграфовой модели. Рассматривается понятие протографа и предлагается его формализация на основе операций ненаправленного прилегания и направленного прилегания элементов. Формализуется понятие архиграфа, рассматривается связь между архиграфом и раскраской протографа, рассматривается операция расширения протографа. Рассматривается связь информационных элементов метаграфа (ИЭМ) с протографовой моделью. Предлагается новый вариант формализации метаграфовой модели на основе точек соединения.

**Ключевые слова:** метаграф, метавершина, протограф, архиграф, информационный элемент метаграфа (ИЭМ).

### Введение

В области инженерии знаний традиционно важную роль играют графовые модели представления знаний. Как ранее, так и в настоящее время, повышенное внимание исследователей в данной области привлекают модели на основе неклассических графовых структур. Если некоторые из таких моделей, в частности, гиперграфы, детально исследованы [Берштейн, 2005], то более сложные модели активно изучаются в настоящее время. В частности, к таким моделям относится метаграфовая модель представления знаний. Будучи изначально предложенной А. Базу и Р. Блэннингом в 2007 году [Basu et al., 2007], модель в дальнейшем получила ряд расширений, независимо предлагаемых различными группами исследователей. В частности, в работе [Гапанюк, 2021] нами была пред-

---

\* Работа выполнена при финансовой поддержке РФФИ (проект №20-07-00770).

ложена аннотируемая метаграфовая модель, но, при этом, недостаточно детально рассматривался вопрос о том, как элементы модели соединяются в единый метаграф. Для ответа на этот вопрос в данной работе будем использовать протографовый подход.

Статья организована следующим образом: в разделе 1 рассматриваются основные положения аннотируемой метаграфовой модели; в разделе 2 рассматривается и формализуется понятие протографа; в разделе 3 рассматривается понятие архиграфа и его связь с раскраской протографа, а также рассматривается операция расширения протографа (архиграфа); в разделе 4 рассматривается «информационный элемент метаграфа» (ИЭМ) и его связь с протографовой моделью; в разделе 5 на основе рассмотренных ранее концепций определяется метаграфовая модель на основе точек соединения.

## 1. Основные положения аннотируемой метаграфовой модели

В данном разделе рассмотрим основные положения аннотируемой метаграфовой модели [Гапанюк, 2021], в дальнейшем модель будет расширена на основе протографового подхода.

Аннотируемая метаграфовая модель расширяет идеи исходной модели А. Базу и Р. Блэннинга [Basu et al., 2007] и идеи работы [Астанин и др., 2012].

В аннотируемой метаграфовой модели метаграф определяется следующим образом:

$$MG = \langle V, MV, E, ME \rangle, \quad (1.1)$$

где  $MG$  – метаграф;  $V$  – множество вершин метаграфа;  $MV$  – множество метавершин метаграфа;  $E$  – множество ребер метаграфа;  $ME$  – множество метаребер метаграфа.

В настоящее время вопрос протографовой формализации метаребер находится в процессе исследования. Поэтому в данной работе метаребра не рассматриваются, а протографовый подход применяется к метавершинам, вершинам и ребрам.

Вершина метаграфа характеризуется множеством атрибутов:

$$v \quad (1.2)$$

где  $v_i$  – вершина метаграфа;  $atr_k$  – атрибут.

Ребро метаграфа характеризуется множеством атрибутов, исходной и конечной вершиной и признаком направленности:

$$, \quad (1.3)$$



где  $e_i$  – ребро метаграфа;  $v_S$  – исходная вершина (метавершина) ребра;  $v_E$  – конечная вершина (метавершина) ребра;  $eo$  – признак направленности ребра ( $eo=true$  – направленное ребро,  $eo=false$  – ненаправленное ребро);  $atr_k$  – атрибут.

Фрагмент метаграфа:

$$, ev_j \in (V \cup E \cup MV), \quad (1.4)$$

где  $MG_i$  – фрагмент метаграфа;  $ev_j$  – элемент, принадлежащий объединению множеств вершин (метавершин) и ребер метаграфа.

Таким образом, фрагмент метаграфа в общем виде может содержать произвольные вершины (метавершины) и ребра.

Метавершина метаграфа:

$$mv_i = \left\{ \{atr_k\}, \{ev_j\} \right\}, \quad mv_i \in MV, \quad ev_j \in (V \cup E \cup MV), \quad (1.5)$$

где  $mv_i$  – вершина метаграфа;  $atr_k$  – атрибут,  $ev_j$  – элемент, принадлежащий объединению множеств вершин (метавершин) и ребер метаграфа. Таким образом, метавершина, в дополнение к свойствам вершины, включает вложенный фрагмент метаграфа.

*Рис. 1. Пример описания метаграфа в аннотируемой метаграфовой модели*

Наличие у метавершин собственных атрибутов и связей с другими вершинами является важной особенностью метаграфов. Это соответствует принципу эмерджентности, то есть приданию понятию нового качества, несводимости понятия к сумме его составных частей. Фактически, как только вводится новое понятие в виде метавершины, оно «получает право» на собственные свойства, связи и т.д., так как в соответствии с принципом эмерджентности новое понятие обладает новым качеством и не может быть сведено к подграфу базовых понятий.

Таким образом, метаграф можно охарактеризовать как «сложный граф с эмерджентностью» или «сложную сеть с эмерджентностью», то есть фрагмент сети, состоящий из вершин и связей, может выступать как отдельное целое.

Пример описания метаграфа показан на рис. 1. Данный метаграф содержит вершины, метавершины и ребра. На рис. 1 показаны три метавершины:  $mv_1$ ,  $mv_2$  и  $mv_3$ . Метавершина  $mv_1$  включает вершины  $v_1$ ,  $v_2$ ,  $v_3$  и связывающие их ребра  $e_1$ ,  $e_2$ ,  $e_3$ . Метавершина  $mv_2$  включает вершины  $v_4$ ,  $v_5$  и связывающее их ребро  $e_6$ . Ребра  $e_4$ ,  $e_5$  являются примерами ребер, соединяющих вершины  $v_2$ - $v_4$  и  $v_3$ - $v_5$ , включенные в различные метавершины  $mv_1$  и  $mv_2$ . Ребро  $e_7$  является примером ребра, соединяющего вершины  $mv_1$  и  $mv_2$ . Ребро  $e_8$  является примером ребра, соединяющего вершину  $v_2$  и метавершину  $mv_2$ . Метавершина  $mv_3$  включает метавершину  $mv_2$ , вершины  $v_2$ ,  $v_3$  и ребро  $e_2$  из метавершины  $mv_1$  а также ребра  $e_4$ ,  $e_5$ ,  $e_8$ , что показывает колоническую структуру метаграфа.

Отметим, что предложенная модель называется аннотируемой, потому что в отличие от [Глоба и др., 2015], в данной модели не выполняется свойство анти-аннотируемости. Одинаковый набор вершин и ребер может быть включен в несколько различных метавершин, которые могут представлять различные ситуации и быть аннотированы различными атрибутами.

Таким образом, текущая версия аннотируемой метаграфовой модели содержит различные элементы модели (вершины, метавершины, ребра). При этом, в текущей версии модели недостаточно детально рассматривается вопрос о том, как элементы модели соединяются в единый метаграф. Для ответа на этот вопрос будем использовать протографовый подход.

## 2. Понятие протографа и его формализация

Понятие протографа вводится в работах [Кручинин, 2017a] и [Кручинин, 2017b]. В соответствии с [Кручинин, 2017b]: «протограф  $P$  задается множеством элементов  $\{p_i\}$ ,  $i = 1, n$  и матрицей соседства (смежности)  $M_{n \times n}$ , состоящей из 0 и 1, где 1 означает соседство (смежность) элемента  $a$  элементу  $b$ ». Таким образом, как отмечается в [Сухобоков, 2022]: «протограф можно рассматривать как граф, не имеющий ребер; роль ребер выполняет прилегание вершин друг к другу».

Отметим, что поскольку определение протографа основано на матрице смежности, то речь идет именно о попарном прилегании элементов протографа друг к другу. Также в [Кручинин, 2017b] отмечается, что прилегание может быть направленным (ориентированным).

На рис. 2 представлен пример неориентированного протографа из четырех элементов, а на рис. 3 пример соответствующего ему ориентированного протографа, направленность между элементами показана стрелками.

Введем формализованные операции прилегания элементов (которые не были предложены в работах [Кручинин, 2017a] и [Кручинин, 2017b]).

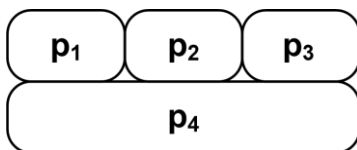


Рис. 2. Пример неориентированного протографа  $P_1$ .

Рис. 3. Пример ориентированного протографа  $P_2$

Для неориентированного протографа операция « $*$ » прилегания элементов протографа может быть определена как двухместная операция:

$$p_1 * p_2. \quad (2.1)$$

Исходя из определения неориентированного протографа, операция прилегания является:

- Коммутативной:  $p_1 * p_2 = p_2 * p_1$ . Порядок объединения двух соседних элементов протографа неважен.

Ассоциативной:  $p_1 * (p_2 * p_3) = (p_1 * p_2) * p_3$ . Порядок объединения нескольких соседних элементов протографа также неважен.

Для ориентированного протографа операция « $**$ » направленного прилегания элементов протографа также может быть определена как двухместная операция (предполагается, что прилегание направлено от левого операнда к правому):

$$p_1 ** p_2. \quad (2.2)$$

Исходя из определения ориентированного протографа, операция направленного прилегания является:

- Некоммутативной:  $p_1 ** p_2 \neq p_2 ** p_1$ . В данном случае выражения  $p_1 ** p_2$  и  $p_2 ** p_1$  обладают различной направленностью.

Ассоциативной:  $p_1 ** (p_2 ** p_3) = (p_1 ** p_2) ** p_3$ . Порядок объединения нескольких соседних элементов протографа в случае сохранения направленности неважен.

На основе введенных операций может быть предложено другое определение протографа:

(2.3)

где  $P$  – протограф;  $PE$  – множество элементов протографа  $\{p_i\}$ ;  $PL^U$  – множество ненаправленных прилеганий элементов протографа  $\{pl_j^U\}$ ;  $PL^D$  – множество направленных прилеганий элементов протографа  $\{pl_k^D\}$ . При этом ненаправленные прилегания задаются с использованием оператора  $*$ , а направленные прилегания с использованием оператора  $**$ .

Тогда, пример неориентированного протографа  $P_1$ , представленный на рис. 2, может быть формально определен следующим образом:

$$P = \langle PE, PL^U, PL^D \rangle,$$

$$PE = \{ p_1, p_2, p_3, p_4 \},$$

$$PL^U = \{ p_1 * p_2, p_2 * p_3, p_1 * p_4, p_2 * p_4, p_3 * p_4 \},$$

$$PL^D = \emptyset.$$

Пример ориентированного протографа  $P_2$ , представленный на рис. 3, может быть формально определен следующим образом:

$$P = \langle PE, PL^U, PL^D \rangle,$$

$$PE = \{ p_1, p_2, p_3, p_4 \},$$

$$PL^U = \{ p_1 * p_4, p_2 * p_4 \},$$

$$PL^D = \{ p_1 ** p_2, p_2 ** p_3, p_3 ** p_4 \}.$$

Таким образом, протографовая модель позволяет отказаться от традиционного понимания графа как совокупности вершин и ребер и рассматривать граф как совокупность прилегающих элементов.

### 3. Архиграф и раскраска протографа

В работах [Кручинин, 2017a] и [Кручинин, 2017b] также вводится понятие архиграфа. В работе [Кручинин, 2017b] дается следующее определение: «Архиграф с числом  $n$  – это протограф с разбиением элементов на  $n$  классов. и соответствующим правилом разбиения. Семейство архиграфов с числом  $n$  – семейство отличных разбиений протографов на  $n$  классов». Каждому классу можно сопоставить цвет, тогда архиграф является протографом, раскрашенным в  $n$  цветов, соответствующих классам.

В работе [Кручинин, 2017b] рассматривается следующий пример: «Простой неориентированный граф является протографом с числом разбиения 2, где элементы 1 класса не могут соседствовать с элементами 1 класса, но каждый элемент 1 класса может иметь сколько угодно отношений соседства с элементами 2 класса».

В работе [Сухобоков, 2022] отмечается, что «аннотируемые метаграфы, описанные в [Гапанюк, 2021], это архиграфы с пятью классами элементов: атрибуты, вершины, метавершины, ребра и метаребра».

Также в [Кручинин, 2017b] вводится операция расширения протографа (расширения архиграфа): «операция расширения заключается в том, что все элементы протографа помечаются одним цветом (то есть относятся к одному классу), но между соседними элементами добавляются элементы второго класса».

На рис. 4 представлен пример протографа  $P_3$ , который является расширением протографа  $P_1$  до двух классов. Элементы второго класса показаны с помощью темного фона, их индексы соответствуют индексам прилегающих элементов первого класса.

*Рис. 4. Пример протографа  $P_3$  (расширение протографа  $P_1$  до двух классов)*

Отметим, что согласно определению [Кручинин, 2017b], элементы второго класса могут прилегать к произвольному количеству элементов первого класса. Элемент  $p_{124}$  является примером элемента второго класса, который прилегает к трем элементам первого класса. Остальные элементы второго класса, представленные на рис. 4, прилегают только к двум элементам первого класса.

Определение одноклассового протографа на основе формулы (2.3) можно расширить до определения архиграфа (многоклассового протографа):

$$AR = \langle P, CLS^n, P.PE \xrightarrow{\varphi} CLS^n \rangle, \quad (3.1)$$

где  $AR$  – архиграф;  $P$  – базовый протограф (в соответствии с формулой (2.3)), на основе которого строится архиграф;  $CLS^n$  – множество классов вершин архиграфа мощностью  $n$ ;  $P.PE \xrightarrow{\varphi} CLS^n$  – отображение

множества вершин  $PE$  базового протографа  $P$  на множество классов  $CLS^n$  (раскраска элементов протографа соответствующими классами из множества  $CLS^n$ ).

Отметим, что операция расширения протографа фактически позволяет добавлять соединительные элементы (второго класса) между основными элементами протографа (первого класса). Этот подход может быть использован для расширения аннотируемой метаграфовой модели.

#### 4. Информационный элемент метаграфа

Отметим, что идеи, отчасти подобные протографовому подходу, высказывались и ранее. В частности, в одной из наших первых работ по метаграфовому подходу [Самохвалов, 2015] была предложена концепция «информационного элемента метаграфа» (ИЭМ), который является элементарной строительной единицей для конструирования метаграфов. В соответствии с [Самохвалов, 2015] ИЭМ определяется как:

(4.1)

где  $id$  – уникальный идентификатор элемента;  $NM$  – наименование элемента;  $VAL$  – значение элемента;  $RL$  – роль элемента;  $lnk_i$  – ссылка на другой ИЭМ;  $atr_j$  – атрибут;  $RL_V$  – роль элемента «вершина»;  $RL_{MV}$  – роль элемента «метавершина»;  $RL_R$  – роль элемента «ребро».

Таким образом, ИЭМ может использоваться для хранения вершины, метавершины и ребра метаграфа. С использованием ссылок  $lnk_i$  реализуются нетипизированные связи между вершиной и ребром, а также связи между метавершиной и входящими в нее элементами.

Фактически, в данном случае ИЭМ изоморфен элементу протографа. Множество ссылок  $lnk_i$  задает нетипизированные связи между соседними ИЭМ аналогично операции ненаправленного прилегания « $\ast$ » в протографах. Роль элемента  $RL$  аналогична раскраске протографа, таким образом в формуле (4.1) задается архиграф из трех классов.

Но ИЭМ соответствует элементу метаграфовой модели (вершине, метавершине, ребру), в то время как связям между ИЭМ в работе [Самохвалов, 2015] не было уделено достаточного внимания, связи задаются с помощью ссылок  $lnk_i$ .

С учетом рассмотренных положений протографового подхода (в особенности операции расширения протографа) и ИЭМ предложим метаграфовую модель, в которой достаточное внимание уделяется как элементам метаграфовой модели, так и связям между ними.

## 5. Метаграфовая модель на основе точек соединения

Введем в метаграфовую модель новый элемент – точку соединения:

$$lp = \langle \text{ИЭМ}_1, \text{ИЭМ}_2, \{atr_k\} \rangle, \quad (5.1)$$

где  $lp$  – точка соединения;  $\text{ИЭМ}_1$  – первый информационный элемент метаграфа, соответствующий выходу точки соединения;  $\text{ИЭМ}_2$  – второй информационный элемент метаграфа, соответствующий выходу точки соединения;  $atr_k$  – атрибут.

Точка соединения объединяет два информационных элемента метаграфа ненаправленным образом, аналогично операции ненаправленного прилегания « $\#$ » в протографах.

На основе определения ИЭМ (4.1) можно отметить, что точка соединения предназначена для замены ссылки  $lnk_i$ . Но принципиальным отличием точки соединения от ссылки  $lnk_i$  является наличие атрибутов у точки соединения. Это открывает значительные возможности для расширения метаграфовой модели. В частности, это позволяет строить нечеткие или вероятностные метаграфы не изменяя при этом самой модели, так как мера нечеткости или вероятность могут быть заданы с помощью атрибутов точки соединения.

Также необходимо отметить, что определение точки соединения (5.1) изоморфно определению ребра метаграфа (1.3) с точностью до признака направленности (параметр  $eo$ ). Но между ними существует принципиальное различие, так как ребро является элементом метаграфовой модели верхнего уровня, а точка соединения является элементом метаграфовой модели нижнего уровня и позволяет «прикреплять» ребра к вершинам и метавершинам.

Базируясь на определении ИЭМ из работы [Самохвалов, 2015] (4.1) и определении точки соединения (5.1), дадим новое определение ИЭМ, учитывающее определение точки соединения:

$$\text{ИЭМ} = \langle id, NM, VAL, RL, \{lp_i\}, \{atr_j\} \rangle. \quad (5.2)$$

Отличие определения (5.2) от определения (4.1) состоит в том, что в определении (5.2) вместо множества связей  $lnk_i$  используется множество точек соединения  $lp_i$ , связанных с данным ИЭМ.

Также в дополнение к множеству элементов  $RL$  из формулы (4.1) введем отношение порядка на элементах этого множества, которые соответствуют элементам метаграфовой модели:

$$RL_{MV} > RL_R > RL_V. \quad (5.3)$$

«Наибольшим» элементом метаграфовой модели является метавершина ( $RL_{MV}$ ). «Наименьшим» элементом метаграфовой модели является вершина ( $RL_V$ ).

Средним элементом является ребро ( $RL_R$ ), так как в соответствии с [Галанюк, 2018] ребро может рассматриваться как частный случай метавершины, содержащий два элемента. Поэтому ребро «меньше» метавершины (так как является частным случаем метавершины), но «больше» обычной вершины, так как содержит две вершины.

Основываясь на введенном отношении порядка (5.3), отметим, что хотя точка соединения не является направленным элементом, в определении явно выделяются два информационных элемента метаграфа ( $ИЭМ_1$  и  $ИЭМ_2$ ), которые она соединяет. Это позволяет формировать точки соединения на основе фиксированных правил, например,  $ИЭМ_1$  всегда соответствует «большемому» элементу метаграфовой модели.

Обобщая введенные понятия, можно дать следующее определение метаграфа:

$$MG = \langle \{ИЭМ_i\}, \{lp_j\} \rangle. \quad (5.4)$$

В соответствии с формулой (5.4) метаграф содержит множество информационных элементов  $ИЭМ_i$  и множество точек соединения  $lp_j$ , связывающих информационные элементы.

Предложенная модель метаграфа на основе точек соединения (5.4) обладает следующими преимуществами:

- Подход на основе ИЭМ позволяет унифицировать метаграфовую модель, представляя элементы метаграфовой модели (вершины, метавершины, ребра) в виде единой структуры ИЭМ с различными параметрами.
- По сравнению с исходной моделью метаграфа, описанной в разделе 1, и подходом на основе ИЭМ (4.1), предложенная модель позволяет гибко соединять элементы метаграфовой модели (вершины, метавершины, ребра) за счет использования точек соединения.
- Использование точек соединения с набором атрибутов позволяет расширять метаграфовую модель, в частности, строить нечеткие или вероятностные метаграфы, не изменяя при этом основные элементы модели (вершины, метавершины, ребра). Например, для точки соединения между вершиной и метавершиной, можно задать атрибут, который будет интерпретироваться как мера нечеткой принадлежности вершины к метавершине.



- В настоящее время активно исследуются вопросы, связанные с хранением метаграфовых данных [Chernenkiy, 2018]. Предложенный подход на основе точек соединения позволит унифицировать способы хранения метаграфовых данных в реляционных, документо-ориентированных и графовых СУБД.

## Заключение

Традиционная метаграфовая модель содержит три вида элементов: вершины, ребра и метавершины. Метавершина, в дополнение к свойствам вершины, включает вложенный фрагмент метаграфа.

Особенностью аннотированной метаграфовой модели является то, что одинаковый набор вершин и ребер может быть включен в несколько различных метавершин, которые могут представлять различные ситуации и быть аннотированы различными атрибутами.

Поскольку в текущей версии метаграфовой модели недостаточно детально рассматривается вопрос о том, как элементы модели соединяются в единый метаграф, то для соединения элементов в данной работе предлагается использовать протографовый подход.

Классический протограф задается множеством элементов и матрицей соседства, в данной работе предлагается альтернативное определение протографа на основании введенных операций ненаправленного прилегания «\*» и направленного прилегания «\*».

Архиграф позволяет разбивать множество элементов протографа на непересекающиеся классы, которые могут быть интерпретированы как цвета раскраски протографа. Операция расширения протографа (архиграфа) позволяет встраивать в существующий архиграф новые классы, задавая прилегание соответствующих элементов классов.

В случае использования подхода на основе «информационных элементов метаграфа» (ИЭМ), ИЭМ изоморфен элементу протографа, а роль ИЭМ (соответствующая вершине, ребру, или метавершине) аналогична раскраске протографа, то есть задает архиграф из трех классов.

Предложенная в данной работе метаграфовая модель на основе точек соединения использует в качестве базовых концепций протографовый (архиграфовый) подход, а также понятие ИЭМ.

Точка соединения позволяет объединить два ИЭМ, задавая для объединения дополнительные атрибуты, которые могут интерпретироваться как мера нечеткости или вероятность вхождения ИЭМ друг в друга.

Таким образом, предложенный подход позволяет расширить метаграфовую модель, строя за счет использования атрибутов точек соединения нечеткие или вероятностные метаграфы, и при этом не изменяя основные элементы модели (вершины, метавершины, ребра).

## Список литературы

- [Астанин и др., 2012] Астанин С.В., Драгныш Н.В., Жуковская Н.К. Вложенные метаграфы как модели сложных объектов // Инженерный вестник Дона. – 2012. – №4. – URL: [ivdon.ru/magazine/archive/n4p2y2012/1434](http://ivdon.ru/magazine/archive/n4p2y2012/1434).
- [Берштейн, 2005] Берштейн Л.С., Боженок А.В. Нечеткие графы и гиперграфы. – М.: Научный мир, 2005. – 256 с.
- [Гапанюк, 2021] Гапанюк Ю.Е. Этапы развития метаграфовой модели данных и знаний // Интегрированные модели и мягкие вычисления в искусственном интеллекте (ИММВ-2021): Сборник научных трудов X-й Международной научно-технической конференции. В 2-х т. Т. 1. – Смоленск, 2021. – С. 190-200.
- [Гапанюк, 2018] Гапанюк Ю.Е. Подход к разработке метаграфового исчисления // Динамика сложных систем – XXI век. – 2018. – Т. 12. № 3. – С. 40-46.
- [Глоба и др., 2015] Глоба Л.С., Терновой М.Ю., Штогрин Е.С. Метаграфы как основа для представления и использования баз нечетких знаний // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS-2015): Материалы V международного науч.-техн. конф. (Минск, 19-21 февраля 2015 года) / редкол.: В.В. Голенков (отв. ред.) [и др.]. – Минск: БГУИР, 2015. – С. 237-240.
- [Кручинин, 2017а] Кручинин С.В. Протографы и архиграфы как обобщение графов // Научно-исследовательские публикации. – 2017. – № 3 (41). – С. 23-33.
- [Кручинин, 2017б] Кручинин С.В. О некоторых обобщениях графов: мультиграфы, гиперграфы, метаграфы, потоковые и портовые графы, протографы, архиграфы // Вопросы науки. – 2017. – № 3. – С. 48-67.
- [Самохвалов, 2015] Самохвалов Э.Н., Ревунков Г.И., Гапанюк Ю.Е. Использование метаграфов для описания семантики и прагматики информационных систем // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. – 2015. – № 1 (100). – С. 83-99.
- [Сухобоков, 2022] Сухобоков А.А. Метаграфово-табличная модель данных для систем управления активами // Искусственный интеллект в автоматизированных системах управления и обработки данных. Сборник статей Всероссийской научной конференции. В 2-х т. Т. 1. – М., 2022. – С. 93-99.
- [Basu et al., 2007] Basu A., Robert W. Blanning. Metagraphs and their applications. – New York: Springer, 2007.
- [Chernenkiy, 2018] Chernenkiy V.M., Gapanjuk Y.E., Kaganov Y.T., Dunin I.V., Lyaskovsky M.A., Larionov V.S. Storing metagraph model in relational, document-oriented, and graph databases // CEUR Workshop Proceedings. – 2018. – 2277. – P. 82-89.

## ВЫБОР ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ ПРЕДОБРАБОТКИ И ТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ НАУЧНЫХ СТАТЕЙ ИЗ ОЗЕРА ДАННЫХ\*

М.М. Гаянова (*gayanova.mm@ugatu.su*)

Е.Ю. Сазонова (*ekaterina\_rassadnikova@mail.ru*)

О.Н. Сметанина (*smoljushka@mail.ru*)

А.К. Сулейманов (*sulaz@mail.ru*)

Уфимский университет науки и технологий, Уфа

В статье рассматриваются вопросы выбора моделей и методов предварительной обработки текстов, представленных научными статьями, автоматически извлеченными из открытых источников и загруженных в озеро данных, их тематического моделирования. Результаты выбора заложены в создаваемое программное решение, которое позволяет провести сбор научных публикаций и хранение их в озере данных, проведение предобработки текста и непосредственно тематическое моделирование с возможностью визуализации результатов. Статья включает современное состояние проблемы, постановку задачи и подход к ее решению, проведение экспериментальных исследований и создание программного решения, в основу которого положены выбранные модели и методы.

**Ключевые слова:** тематическое моделирование, озеро данных, методы предварительной обработки текстов, векторизация текста.

### Введение

В настоящее время в связи с существенным увеличением объемов электронных документов возникает необходимость использования методов и инструментальных средств для автоматической организации, поиска, индексирования и сканирования больших коллекций документов. При разработке высокотехнологичных решений, критических систем требуется постоянный мониторинг научных статей, большая часть которых опубли-

---

\* Работа выполнена при финансовой поддержке РФ (проект № 22-19-00471).

ликована в электронных журналах. «Ручной» поиск в этом случае является трудоемкой задачей. Поэтому актуальной будет разработка инструментальных решений для автоматического сбора, хранения и определения тематики текстовых документов. Последнее может быть получено с помощью методов тематического моделирования (ТМ), представляющего собой построение коллекции текстовых документов, и позволяющего определить, к каким темам относится каждый из них [Булатов, 2020]. Тематическая модель позволяет для каждого документа найти темы, которые его описывают, а также показывает, какие слова характеризуют ту или иную тему.

ТМ находит применение в различных приложениях: от категоризации, аннотирования текстов, до анализа сообщений из социальных сетей. Традиционно ТМ используется для работы с естественным языком, в биоинформатике (например, для анализа аминокислотных последовательностей), медицине (например, поиск по веб-страницам медицинской тематики с учетом персонализации вопросов), компьютерном зрении (например, обнаружение визуальных шаблонов (тем)) и пр.

Вопросам ТМ посвятили свои работы Воронцов К. В., Коршунов А., Гомзин А., Deerwester S., Dumais S. T., Furnas G. W., Landauer Th. K., Harshman R., Hofmann Th., Blei D. M., Jordan M., Griffiths T. L., Steyvers M. и многие другие.

В статье рассматривается проблемная ситуация, связанная с необходимостью разработки моделей и методов тематического моделирования научных текстов. В первом разделе рассматривается описание современного состояния проблемы с анализом существующих программных решений, во втором разделе – постановка задачи ТМ и подход к ее решению, в третьем разделе – проведение экспериментальных исследований и интерпретация полученных результатов.

## **1. Современное состояние проблемы**

Вопросами ТМ с разработкой различных моделей исследователи начали заниматься в конце прошлого столетия (рис. 1) [Papadimitriou, 1998], [Hofmann, 1999], [Blei et al., 2003]. При анализе текстов на естественном языке (ЕЯ) тематическое моделирование позволяет проводить обнаружение скрытых тематических шаблонов, присутствующих в коллекции; аннотирование документов по этим темам и использование аннотаций для систематизации, поиска и обобщения текстов.

*Рис. 1. Истоки методов тематического моделирования*

Первые модели семантического моделирования получили дальнейшее развитие. ТМ рассматривают [Гринева и др., 2009] как метод поиска группы слов (например, темы) из коллекции документов, которая наилучшим образом представляет информацию в коллекции, или как форму интеллектуального анализа текста (способ получения повторяющихся шаблонов слов в текстовом материале). Современные тематические модели, как правило, являются расширением латентного размещения Дирихле (Latent Dirichlet allocation, LDA), как одного из самых популярных методов ТМ, имеющего целью нахождение тем, к которым принадлежит документ, на основе содержащихся в нем слов. Метод основывается на предположениях: каждый документ может быть описан некоторым набором тем; каждая тема может быть описана некоторым набором слов.

Латентный (скрытый) семантический анализ (Latent semantic analysis, LSA) предполагает, что близкие по значению слова встречаются в похожих частях текста (гипотеза распределения). Матрица, содержащая количество слов в документе (строки представляют уникальные слова, а столбцы представляют каждый документ), строится из большого фрагмента текста, а математический метод, называемый разложением по сингулярным значениям (Singular Value Decomposition, SVD), используется для уменьшения количества строк при сохранении структуры сходства среди столбцов. Документы сравниваются путем определения косинуса угла между двумя векторами (или скалярного произведения между нормализацией двух векторов), образованного любыми двумя столбцами. Значения, близкие к 1, представляют очень похожие документы, тогда как значения, близкие к 0, представляют очень разные документы [Dumais, 2005].

Вероятностный латентно-семантический анализ, также известный как вероятностное латентно-семантическое индексирование, базируются на статистическом методе анализа корреляции двух типов данных, являясь дальнейшим развитием латентно-семантического анализа. По сравнению с обычным латентно-семантическим анализом, основанным на линейной алгебре, представляя способ снижения размерности матрицы (разложение диагональной

матрицы по сингулярным значениям), вероятностный латентно-семантический анализ основан на смешанном разложении, в свою очередь берущим свое начало из модели скрытых классов. Данный подход более принципиален, поскольку имеет прочную основу в области статистики.

В работе [Коршунов и др., 2012] Коршунов А. и Гомзин А. рассматривают различные подходы тематического моделирования: латентно-семантическое индексирование, вероятностные тематические модели, Байесовские сети, скрытое размещение Дирихле. Авторы описывают методы оценивания качества результатов, приводят примеры существующих программных реализаций.

В статье [Котов и др., 2005] Котов Э. М. и Целых А.Н. разделяют всю совокупность методов тематического анализа текста на две группы: статистический анализ (заключается в подсчете количества повторений слов в тексте и использовании результатов подсчета для конкретных целей) и лингвистический анализ (подразделяется на четыре взаимодополняющие группы (лексический анализ, морфологический анализ, синтаксический анализ, семантический анализ)).

Авторы статьи [Воронцов и др., 2012] предлагают обобщенное семейство вероятностных тематических моделей коллекций текстовых документов, в котором эвристики регуляризации, сэмплирования, частого обновления параметров, робастности относительно шума и фона могут включаться независимо друг от друга в любых сочетаниях, порождая как известные модели PLSA (Probabilistic Latent Semantic Analysis), LDA, CVB0 (Collapsed Variational Bayes), SWB (Special Words with Background), так и новые.

В статье [Blei et al., 2009] описывается скрытое распределение Дирихле и его применение для декомпозиции и исследования большой коллекции документов. Как утверждают авторы, ТМ может предоставить полезное представление о большой коллекции с точки зрения коллекции в целом, отдельных документов и взаимосвязей между документами.

В статье [Griffiths et al, 2004] авторы представили алгоритм статистического вывода для скрытого распределения Дирихле, генеративную модель для документов, в которой каждый документ рассматривается как смесь тем, и показали, как этот алгоритм можно использовать для понимания содержания научных документов.

В работе [Liu et al., 2016] описывается применение ТМ в сфере биоинформатики, в отличие от традиционных средств обработки данных в биоинформатике. Предлагаемое решение расширяет возможности исследователей интерпретировать биологическую информацию. Также авторами отмечается отсутствие тематических моделей, оптимизированных для конкретных биологических данных.

Автор статьи [Rosen-Zvi, 2004] рассматривает авторско-тематическую модель для ТМ, которая расширяет латентное размещение Дирихле, с помощью включения информации об авторстве. Эта модель обеспечивает значительно улучшенную предсказательную силу по сравнению с более бедной авторской моделью.

В работе [Zhijun, 2011] исследуется проблема обнаружения и сравнения географических тем из документов, связанных с GPS (Global Positioning System). Связанные с GPS документы становятся популярными с распространением технологий определения местоположения. Авторами предлагаются и сравниваются три способа моделирования географических тем: модель, управляемая местоположением; модель, управляемая текстом; новая совместная модель под названием LGTA (Latent Geographical Topic Analysis – скрытый географический анализ темы), которая объединяет местоположение и текст.

В [Manikajun, 2019] было проанализировано 928 полнотекстовых исследовательских статей, извлеченных из журнала DESIDOC Journal of Library and Information Technology за период 1981–2018 гг. с использованием скрытого распределения Дирихле. За 38 лет было выявлено 50 основных тем, из которых только 26 тем были уникальными по своему характеру. Как утверждает автор, ТМ не только помогает исследователю определить актуальные темы или связанные области в соответствии с их областью интересов, но также помогает им определять новые концепции и области с течением времени.

В статье [Yuan et al., 2021] идет речь о нейронных тематических моделях (Neural Topic Model, NTM), которые применяют глубокие нейронные сети к ТМ. Несмотря на успех, NTM обычно игнорируют два важных аспекта: для обучения используется только информация о количестве слов на уровне документа, и внешние семантические знания о документах, предложениях и словах не используются для обучения. Для решения этих проблем авторы предлагают модель NTM с вариационным автокодировщиком (Variational Autoencoder, VAE), которая совместно восстанавливает количество слов в предложениях и документах, используя комбинации тематических встраиваний из набора слов (Bag of words) и предварительного обученных семантических встраиваний.

*Анализ существующих программных решений. Stanford Topic Modeling Toolbox.* Стэнфордский набор инструментов тематического моделирования (ТМТ) [STMT, 2022] может быть использован социологами, которые хотят выполнять анализ наборов данных, содержащих значительный текстовый компонент. В набор включены инструменты с функционалом: импорт и управление текстом из ячеек в Excel и других электронных таблиц; обучение тематических моделей (LDA, размеченный LDA и PLDA) для

создания описания текста; выбор параметров (например, количество тем) с помощью процесса, управляемого данными; создание многофункциональных выходных данных, совместимых с Excel, для отслеживания использования слов по темам, времени и другим группам данных.

*Gensim* как библиотека с открытым исходным кодом для неконтролируемого тематического моделирования и обработки естественного языка с использованием современного статистического машинного обучения, реализован на Python и Cython для повышения производительности. Данный фреймворк разработан для обработки больших текстовых коллекций с использованием потоковых данных и инкрементных онлайн-алгоритмов, что отличает его от большинства других программных пакетов машинного обучения, ориентированных только на обработку в памяти.

В качестве основных возможностей следует выделить: независимость алгоритмов от памяти, (возможность обрабатывать ввод больший, чем ОЗУ); интуитивно понятные интерфейсы; простое подключение своего собственного входного корпуса / поток данных (тривиальный потоковый API (Application Programming Interface)); легкое расширение с помощью других алгоритмов векторного пространства (API тривиального преобразования); эффективные многоядерные реализации популярных алгоритмов, таких как скрытый семантический анализ онлайн (LSA / LSI (Latent Semantic Indexing) / SVD), скрытое распределение Дирихле (LDA), случайные проекции (RP), иерархический процесс Дирихле (Hierarchical Dirichlet Process, HDP) или глубокое обучение word2vec; распределенные вычисления (может запускать скрытый семантический анализ и скрытое распределение Дирихле на кластере компьютеров).

*Familia* [Jiang et al, 2018] представляет собой проект с открытым исходным кодом, который реализует три популярные тематические модели, основанные на крупномасштабных промышленных данных: латентное размещение Дирихле (LDA), Sentence LDA и Topical Word Embedding (TWE). Кроме того, *Familia* предлагает несколько инструментов, включая *lda-infer* и *lda-query-doc-sim*. Данный инструмент можно применить для решения многих задач, таких как классификация документов, кластеризация документов и персональные рекомендации. *Familia* предлагает два метода вычисления семантического сходства между документами: семантическое сходство между короткими и длинными документами, используемое для извлечения ключевых слов и вычисления семантического сходства запроса и документа; семантическое сходство между длинными документами, применяемое для вычисления семантического сходства между профилем пользователя и новостной статьей.



Программная платформа NLTK (Natural Language Toolkit) [Фаткулин, 2015] используется для ТМ и классификации текста, предоставляя множество корпусных и лексических ресурсов. Указанные ресурсы можно использовать для обучения моделей, для обработки текста, включая токенизацию, выделение корней, тегов, синтаксический и семантический анализы.

Сервис обработки естественного языка *Amazon Comprehend* использует машинное обучение для поиска взаимосвязей в тексте. Сервис определяет язык текста; извлекает ключевые фразы, места, людей, бренды или события; понимает, насколько текст положительный или отрицательный; анализирует текст с помощью токенизации и частей речи; и автоматически организует коллекцию текстовых файлов по темам.

Результаты анализа существующих программных решений в области ТМ показали, что они работают с готовыми коллекциями данных. Авторы статьи в своей работе рассматривают задачу немного шире: от автоматического извлечения научных публикаций, находящихся в открытом доступе, размещения коллекции документов в хранилище данных, их тематического моделирования и поиска в хранилище схожих по требуемой тематике документов. Для этого также проведен анализ решений для хранения больших объемов данных.

*Хранение данных в озере данных.* Озеро данных (DataLake) позволяет хранить большие объемы информации в структурированном, слабоструктурированном и неструктурированном виде, являясь удобным решением для компаний, в которых скапливаются большие объемы информации, например, данные датчиков, лог-файлы, социальные данные.

В отличие от иерархического хранилища данных, в котором данные хранятся в файлах и папках, озеро данных имеет плоскую архитектуру [Data lake, 2022]. Каждый элемент данных в озере имеет уникальный идентификатор и помечается набором метаданных. Примерами озер данных являются Apache Hadoop, Google Cloud Storage, Amazon S3, Microsoft Azure DataLake.

## **2. Постановка задачи тематического моделирования и подход к ее решению**

Решаемая задача предполагает разработку программного решения, позволяющего осуществлять автоматическое извлечение научных публикаций, находящихся в открытом доступе, размещение собранной коллекции документов в хранилище данных, их тематическое моделирование и поиск в хранилище схожих по требуемой тематике документов.

Общая постановка задачи выбора инструментальных средств для решения может быть представлена следующим образом.

Дано: множество исходных документов  $A$ ; множество моделей тематического моделирования  $F$ ; где отображение  $f: a \rightarrow b$ , обеспечивает соответствие между элементами множеств  $A$  и  $B$  ( $b$  - элемент некоторого множества  $B$  (тема)). Элементы множества  $f$  характеризуются метрикой  $s$ .

Выявить: модель  $f'$  с учетом результатов проведенных экспериментов.

Математическая постановка задачи, касающаяся собранных документов, может быть сформулирована следующим образом. Дано:  $W$  - словарь терминов,  $D$  - коллекция документов,  $n_{dw}$  - число вхождений термина  $w$  в документе  $d$ ,  $n_d$  - длина документа  $d$ . Документ  $d$  представляет собой смесь распределений  $p(w|t)$  с весами  $p(t|d)$ .

Необходимо найти: параметры вероятностной тематической модели по формуле:

$$p(w|d) = \sum_{t \in T} p(w|t)p(t|d),$$

где  $p(w|t)$  - вероятности терминов  $w$  в каждой теме  $t$ ,  $p(t|d)$  - вероятности тем  $t$  в каждом документе  $d$ .

Задача ТМ решается с помощью вероятностных моделей с учетом ряда базовых предположений (рис. 2).

*Рис. 2. Базовые предположения в задаче тематического моделирования*

Обобщенный алгоритм решения задачи определения тематики научной статьи состоит из следующих этапов (рис. 3):

*Предобработка текста.* Перед тем как подать некоторый текст в модель определения тематики, нужно произвести предварительную обработку данных. Это связано с тем, что в исходных текстах есть много «лиш-

ней» информации, которая не несет важной смысловой нагрузки. Это, например, числа, знаки препинания, часто употребляемые слова – вводные слова или союзы и предлоги. Также нужно разбить предложения на отдельные слова (токенизация) и привести слова к нормальной форме (лемматизация).

Процесс предобработки состоит из следующих этапов: токенизация (исходные предложения в тексте разбиваются на отдельные слова – токены) может быть выполнена с помощью методов, представленных в библиотеке `gensim` (`gensim.utils.simple_preprocess`.); исключение стоп-слов (с помощью списка стоп-слов нужно пройти по тексту и исключить эти слова) можно выполнить с помощью `gensim`; лемматизация (слова приводятся к нормальной форме) можно использовать библиотеку `spaCy` [spaCy, 2022]. Для этого потребуется дополнительно скачать данные корпуса «`ru_core_news_sm`».

*Рис. 3. Этапы обобщенного алгоритма решения задачи определения тематики*

*Выбор методики векторного представления текста.* Для определения тематики текста необходимо передать этот текст некоторой модели (например, модели машинного обучения), которая способна определять тематику текста. Перед этим нужно представить текст в векторной форме, так как модели не принимают данные для анализа в текстовой форме. Для этих целей существуют специальные подходы моделирования языка, которые направлены на то, чтобы сопоставить словам (или фразам) некоторый вектор из словаря векторов небольшой размерности [Федоренко и др., 2018].

Метод «Bag of Words» (BoW) является наиболее простым способом представления текста в векторном виде; в нем текст представляется в виде «мешка слов» без учета положения их в тексте и грамматики. Сохраняется лишь количество каждого слова из текста. Метод используется в обработке естественного языка и информационном поиске [Нугуманова и др., 2016]. Для повышения эффективности метода за счет учета порядка следования слов часто применяют биграммы, триграммы и  $n$ -граммы.

Мера TF-IDF позволяет оценить важность слова в документе, и, как следствие, обеспечить для часто встречающихся слов, но не несущих полезной смысловой информации (например, предлоги и союзы), снижение их важности. Формулы для расчета меры TF-IDF [Михайлов и др., 2015]:

$$tf(t, d) = \frac{n_t}{\sum_k n_k},$$

где  $n_t$  – число вхождений слова  $t$  в документ  $d$ ,  $\sum_k n_k$  – общее количество слов в документе.

$$idf(t, D) = \log \frac{|D|}{|\{d_i \in D \mid t \in d_i\}|},$$

где  $|D|$  – количество документов в коллекции,  $|\{d_i \in D \mid t \in d_i\}|$  – число документов из коллекции  $D$ , в которых встречается  $t$ .

Основание логарифма не имеет значения, так как изменение основания ведет к изменению веса каждого слова на постоянный множитель.

TF-IDF рассчитывается по формуле:

$$tf - idf(t, d, D) = tf(t, d) \cdot idf(t, d, D).$$

*Выбор метода определения тематики текста.* Метод LSA, базирующийся на разложении матрицы документов и терминов на отдельные матрицы «документ-тема» и «тема-термин». Первым шагом является создание матрицы «документ-термин». Каждая строка представляет документ, а каждый столбец представляет слово. В простейшей версии LSA можно использовать простой подсчет слов в документе. На практике, однако, простые подсчеты не показывают хороших результатов, потому что они не учитывают значение каждого слова в документе.

Построенная в результате матрица обычно получается достаточно большой и разреженной. Для оптимизации применяют метод SVD, позволяющий разложить исходную матрицу на множество ортогональных мат-

риц, линейная комбинация которых является достаточно точным приближением к исходной [Сафонова, 2016]. Разложение матрицы осуществляется по формуле:

$$DW_{n \times k} = DT_{n \times t} \cdot TW_{t \times k},$$

где  $n$  – количество документов,  $k$  – количество слов в словаре,  $t$  – количество тем.

*Латентное размещение Дирихле*, позволяя выявить скрытые семантические структуры в коллекции текстовых документов, может быть описано графической моделью (рис. 4).

Метод LDA основан на вероятностной модели:

$$p(d, w) = \sum_{t \in T} p(d) \cdot p(w|t) \cdot p(t, d),$$

где  $d$  – документ,  $w$  – слово,  $t$  – тема,  $T$  – множество тем,  $p(d)$  – априорное распределение на множестве документов,  $p(w|t)$  – условное распределение слова  $w$  в теме  $t$ ,  $p(t|d)$  – условное распределение темы  $t$  в документе  $d$ .

*Рис. 4. Модель LDA:  $D$  – количество документов,  $N$  – количество слов в заданном документе,  $\alpha$  – распределение Дирихле тем по каждому документу,  $\beta$  – распределение Дирихле слов по темам,  $\theta_i$  – распределение тем для  $i$ -го документа,  $\phi_k$  – распределение слов в теме  $k$ ,  $z_{ij}$  – тема для  $i$ -го слова в документе,  $w_{ij}$  – конкретное слово*

Для подбора параметров модели LDA может использоваться семплирование по Гиббсу [Koltcov et al., 2016], которое генерирует выборки распределений для множества случайных величин. Данный метод носит итеративный характер. В начале алгоритма берется некоторое начальное значение. Затем на каждой итерации алгоритм берет случайную величину и выбирает ее значение при условии фиксированных остальных случайных величин.

*Меры оценки качества тематической модели.* Оценки качества подразделяются на внутренние и внешние критерии. Внутренние позволяют оценить качество модели по выходным данным (матрицам). Внешние критерии определяют качество модели в зависимости от того, как модель решает конкретную прикладную задачу. Ввиду сложности реализации внешней оценки моделей, в данном исследовании предложено рассмотреть внутренние оценки: перплексия, демонстрирующая, насколько хорошо распределение предсказывает выборку (меньшие значения перплексии говорят о лучшем качестве модели); при экспертной оценке качество модели определяется группой экспертов путем соответствия тем, к которым отнесены тексты из коллекции; когерентность показывает, насколько часто слова, встречающиеся рядом в текстах, оказываются в списке наиболее часто встречающихся слов заданной темы. В результате анализа оценок выбрана когерентность ввиду того, что: ее вычисление может быть произведено в автоматическом режиме; не требуется привлечения экспертов; она наиболее схожа с оценкой реальных людей при интерпретации тем в документах; плохой интерпретируемостью результатов оценки перплексии.

Существуют различные методы для измерения когерентности [Karadia, 2019]: *мера*  $C_v$  основана на скользящем окне, сегментировании верхних слов по одному набору и мере косвенного подтверждения, использующей нормализованную точечную взаимную информацию (NPMI) и косинусное сходство; *мера*  $C_p$  основана на скользящем окне, сегментации верхних слов на один предшествующий и мере подтверждения согласованности Фительсона; *мера*  $C_{uci}$  основана на скользящем окне и поточечной взаимной информации (PMI) всех пар слов заданных верхних слов; *мера*  $C_{umass}$  основана на количестве совпадений документов, сегментации с одним предшествующим и логарифмической условной вероятности в качестве меры подтверждения; *мера*  $C_{nprmi}$  представляет собой расширенную версию когерентности  $C_{uci}$ , использующую нормализованную точечную взаимную информацию (NPMI); *мера*  $C_a$  основана на контекстном окне, попарном сравнении верхних слов и косвенной мере подтверждения, которая использует нормализованную точечную взаимную информацию (NPMI) и косинусное сходство.

*Сбор текстов из открытых источников.* Сбор научных статей осуществляется путем парсинга сайта КиберЛенинка [CyberLeninka, 2022], который представляет собой научную электронную библиотеку, построенную на парадигме открытой науки (Open Science), основными задачами которой является популяризация науки и научной деятельности. На сегодняшний день в библиотеке представлено 2 618 975 научных статей.

*Алгоритм сбора статей с сайта* (рис. 5) предусматривает следующую последовательность действий: запускается внешний цикл по всем категориям статей (математика, физика, компьютерные науки и т.д.), для каждой категории осуществляется получение списка страниц с текстами научных статей; запускается цикл по всем страницам, для каждой страницы идет сбор текстов со страницы и добавление в список статей для данной категории; по завершению цикла по страницам, осуществляется сохранение списка текстов категории в озеро данных; переход к следующей категории и процесс повторяется.

*Процесс поиска похожих статей.* Для выполнения поиска похожих статей необходимо иметь размеченную выборку данных. Для разметки были использованы статьи, собранные из открытых источников.

Процесс состоит из следующих этапов: разметка собранных статей с помощью тематической модели; сохранение результатов моделирования в БД (в качестве результатов модель выдает темы, обнаруженные в документе, а также ключевые слова для каждой темы); выбор наиболее вероятной темы (модель также выдает вероятности для каждой из тем; большие значения вероятности свидетельствуют о более релевантной теме; в процессе разметки выбирается тема с самой высокой вероятностью); выбор ключевых слов (по каждой теме выбирается топ 7 слов с наиболее высокими весами; выбранные слова будут характеризовать данный документ).

*Процесс мониторинга научных публикаций.* Процесс состоит из следующих этапов: подключение к сайту со статьями; поиск статей, которые удовлетворяют заданным темам. Каждая статья анализируется с помощью тематической модели и в случае наличия заданных тем, сохраняется в БД и добавляется в результаты мониторинга; по завершению мониторинга, обновляется статус мониторинга на «завершен».

В результате сравнения существующих сервисов для организации озера данных была выбрана платформа ApacheNadoop, ввиду возможности свободного неограниченного использования и разворачивания на собственном компьютере.



*Рис. 5. Алгоритм сбора данных*



### 3. Проведение эксперимента и интерпретация результатов

Подготовка обучающего набора данных. Для сбора данных был написан парсер на языке программирования C#. Для работы с DOM-представлением страницы была использована библиотека AngleSharp [AngleSharp, 2022]. Статьи были извлечены из следующих категорий: медицинские науки; гуманитарные науки; естественные и точные науки; сельскохозяйственные науки; техника и технологии; социальные науки. Для каждой категории было собрано примерно по 1000 статей. Количество статей по каждой категории приведено в табл. 1.

Таблица 1

Категория	Количество статей
медицинские науки	1021
гуманитарные науки	1011
естественные и точные науки	1000
сельскохозяйственные науки	1008
техника и технологии	1044
социальные науки	1123

В табл. 2 приведены примеры названий статей для каждой категории. Во время сбора статей также проводилась частичная предобработка текста. Удалялись лишние переводы строк, обрезалась часть текста в начале документа, которая принадлежит предыдущей статье.

Таблица 2

Категория	Название	Авторы
медицинские науки	Изменение функции гипофиз-гонадной системы у больных хорионкарциномой матки под влиянием аутогемохимии терапии	Иванова В.А., Левченко Н.Е., Вереникина Е.В., Кучерявая О.Г.
гуманитарные науки	Морально-психологическое состояние личного состава: сущность и диагностика	Балаян А.И.
естественные и точные науки	Принципы проектирования интегральной модели оценки надежности информационно-вычислительных систем	Гузик В.Ф., Самойленко А.П., Мунтян Е.Р.
Сельскохозяйственные науки	Экономическая эффективность развития лабораторного животноводства в филиале «Андреевка» ГУ НЦБМТ РАМН	Матвеевко Е.Л., Дубакина Л.Г., Афонин К.В.
техника и технологии	Математическое моделирование пространственно-неоднородной задачи биологической кинетики применительно к акватории Таганрогского залива	Никитина А.В.

В ходе эксперимента нужно выбрать способ векторного представления BoW или TF-IDF) и проверить необходимость использования  $n$ -грамм.

Также нужно выбрать метод для построения тематической модели (LSA, LDA).

Для выбора оптимального количества тем был осуществлен расчет когерентности для различного количества тем (от 6 до 19) (табл. 3).

Таблица 3

Количество тем	Когерентность	Количество тем	Когерентность
6	0,469	13	0,42
7	0,469	14	0,44
8	0,482	15	0,43
9	0,501	16	0,4
10	0,493	17	0,39
11	0,469	18	0,38
12	0,443	19	0,4

Наибольшее значение когерентности достигается для 9 тем.

*Результаты эксперимента для модели LDA.* Результаты построения тематических моделей на основе латентного размещения Дирихле для векторных представлений BoW и TF-IDF представлены в табл. 4. В таблицах представлены параметры модели, на которых были достигнуты наиболее высокие значения когерентности.

Таблица 4

Кол-во тем	Alpha		Beta		Когерентность	
	BoW	TF-IDF с использованием биграмм и триграмм	BoW	TF-IDF с использованием биграмм и триграмм	BoW	TF-IDF с использованием биграмм и триграмм
6	0,01	0,1	0,61	0,31	0,545	0,580
7	asymmetric	0,1	0,9	symmetric	0,570	0,521
8	0,9	0,1	symmetric	0,31	0,563	0,579
9	0,9	0,1	0,61	0,31	0,579	0,573

Таким образом, наиболее высокое значение когерентности (0,58) для модели LDA получено с помощью метода векторизации BoW.

*Результаты эксперимента для модели LSA.* Результаты построения тематических моделей на основе латентного семантического анализа для векторных представлений VoW и TF-IDF представлены в табл. 5.

Таблица 5

Кол-во тем	Когерентность	
	VoW	TF-IDF с использованием биграмм и триграмм
6	0,482	6
7	0,475	7
8	0,466	8
9	0,452	9

По результатам эксперимента для модели LSA, было выяснено, что наиболее высокое значение когерентности (0,531) получено с помощью векторного представления TF-IDF с использованием биграмм и триграмм. Список тем, полученных с помощью данной модели включает номер темы и список слов, которые представляют тему:

(0,

'0.012\*"животное" + 0.007\*"собака" + 0.007\*"исследование" + 0.004\*"случай" + 0.004\*"рис" + 0.003\*"результат" + 0.003\*"вид" + 0.003\*"день" + 0.002\*"кошка" + 0.002\*"теленок"),

(1,

'0.006\*"информация" + 0.006\*"социальный" + 0.006\*"человек" + 0.006\*"информационный" + 0.004\*"новый" + 0.004\*"культура" + 0.004\*"являться" + 0.004\*"современный" + 0.004\*"образ" + 0.004\*"система"),

(2,

'0.010\*"группа" + 0.008\*"клетка" + 0.007\*"исследование" + 0.007\*"уровень" + 0.006\*"препарат" + 0.005\*"показатель" + 0.005\*"активность" + 0.005\*"больных" + 0.005\*"результат" + 0.005\*"метод"),

(3,

'0.010\*"год" + 0.004\*"время" + 0.004\*"русский" + 0.004\*"первый" + 0.003\*"новый" + 0.003\*"статья" + 0.003\*"военный" + 0.003\*"история" + 0.003\*"советский" + 0.003\*"государственный"),

(4,

'0.014\*"шрифт" + 0.012\*"труба" + 0.004\*"производство" +  
0.003\*"тоннель" + 0.003\*"трубопровод" + 0.003\*"логотип" +  
0.003\*"шрифтовой" + 0.002\*"семиотический\_базис" +  
0.002\*"визуальный\_навигация" + 0.002\*"выпуск")',

(5,

'0.013\*"система" + 0.006\*"работа" + 0.005\*"модель" +  
0.005\*"значение" + 0.005\*"являться" + 0.004\*"задача" + 0.004\*"время" +  
0.004\*"метод" + 0.004\*"оценка" + 0.004\*"результат")].

Число перед словом показывает его вес в теме.

Можно сопоставить темы, полученные с помощью тематической модели с категориями научных статей, собранных из Интернета (табл. 6):

Таблица 6

Номер темы, полученной с помощью тематической модели	Категория статьи
0	сельскохозяйственные науки
1	социальные науки
2	медицинские науки
3	гуманитарные науки
4	техника и технологии
5	естественные и точные науки

Тематическая модель смогла определить все категории статей, которые были собраны из сети Интернет. Таким образом, можно сделать вывод, что тематическая модель хорошо решает прикладную задачу.

По итогам вычислительного эксперимента обнаружено, что наиболее качественной является тематическая модель на основе латентного размещения Дирихле с использованием векторного представления TF-IDF. Данная модель будет использована в программном обеспечении.

#### **4. Визуализация результатов тематического моделирования**

С помощью библиотеки pyLDavis [pyLDavis, 2022] можно представить результаты тематического представления в удобном для восприятия виде. Библиотека извлекает информацию из тематической модели для создания интерактивной веб-визуализации (рис. 6). Кругами выделены темы. Круги не пересекаются, что свидетельствует о хорошем качестве тематической модели.

*Рис. 6. Графический вид результатов тематического моделирования*

При нажатии на тему справа отображается список наиболее частых слов данной темы (рис. 7).

*Рис. 7. Список слов, описывающих тему 1*

## Заключение

Анализ современного состояния проблемы позволил выбрать комплекс эффективных моделей и методов для проведения предварительной обработки данных, векторизации текста, тематического моделирования, а также озеро данных для хранения собранных с помощью парсера научных статей из открытых источников, в частности, с сайта КиберЛенинка. Для сбора данных написан парсер на языке программирования C#. Для работы с DOM-представлением страницы использована библиотека AngleSharp.

Экспериментальные исследования позволили выбрать модель (тематическая модель на основе латентного размещения Дирихле) и метод (TF-IDF), которые использованы для анализа данных в создаваемом программном решении. Для визуализации результатов использована библиотека pyLDAvis.

Статьи извлечены из ряда категорий: медицинские науки; гуманитарные науки; естественные и точные науки; сельскохозяйственные науки; техника и технологии; социальные науки. Полученное решение может быть использовано и для отдельной категории, например, медицинские науки.

Созданное программное решение позволяет осуществлять мониторинг научных статей, удовлетворяющих заданным темам.

## Список литературы

- [**AngleSharp, 2022**] AngleSharp – библиотека.NET [Электронный ресурс]. – URL: <https://anglesharp.github.io/> (дата обращения: 18.09.2022).
- [**Blei et al., 2003**] Blei D., Ng A., Jordan M. Latent Dirichlet allocation // Journal of Machine Learning Research. – 2002. – No. 3. – Doi:10.1162/jmlr.2003.3.4-5.993.
- [**Blei et al., 2009**] Blei D., Lafferty J. Topic Models. Text Mining: Classification, Clustering, and Applications // Chapman & Hall: CRC Data Mining and Knowledge Discovery Series. – 2009.
- [**CyberLeninka, 2022**] Научная электронная библиотека [Электронный ресурс]. – URL: <https://cyberleninka.ru/> (дата обращения: 17.09.2022).
- [**Data Lake, 2022**] Simple English Wikipedia [Электронный ресурс]. – URL: [https://simple.wikipedia.org/wiki/Data\\_lake#](https://simple.wikipedia.org/wiki/Data_lake#) (дата обращения: 18.09.2022).
- [**Dumais, 2005**] Dumais S. Latent Semantic Analysis // Annual Review of Information Science and Technology. – No. 38. – Doi:10.1002/aris.1440380105.
- [**Griffiths et al, 2004**] Griffiths T., Steyvers M. Finding scientific topics // Proc. of the National Academy of Sciences, USA. 2004. – Doi:10.1073/pnas.0307752101. PMID 14872004.
- [**Hofmann, 1999**] Hofmann T. Probabilistic Latent Semantic Indexing // Proc. of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval.

- [Jiang et al, 2018]** Jiang D., Song Y., Lian R., Bao S., Peng J., He H., Wu H. Familia: A Configurable Topic Modeling Framework for Industrial Text Engineering. [Электронный ресурс]. arXiv. Служба распространения и архив с открытым доступом. – URL: <https://arxiv.org/pdf/1808.03733.pdf> (дата обращения 04.09.2022).
- [Kapadia, 2019]** Shashank Kapadia. Evaluate Topic Models: Latent Dirichlet Allocation (LDA) [Электронный ресурс]. Towards Data Science. – 2019. – URL: <https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0> (дата обращения: 17.09.2022).
- [Koltcov et al., 2016]** Koltcov S.N., Nikolenko S.I., Koltsova E.Y. Gibbs Sampler Optimization for Analysis of a Granulated Medium // Journal of Technical Physics Letters. – 2016. – Vol. 42, No. 16.
- [Liu et al., 2016]** Liu L.; Tang L.; et al. An overview of topic modeling and its current applications in bioinformatics // SpringerPlus. – 2016. – No. 5 (1). – Doi: 10.1186/s40064-016-3252-8.
- [Manikajun, 2019]** Manikajun L. Mapping of topics in DESIDOC Journal of Library and Information Technology, India: a study // Journal of the Scientometrics. – 2019. – No. 120 (2). – Doi: 10.1007/s11192-019-03137-5. – ISSN 0138-9130. S2CID 174802673.
- [Papadimitriou et al., 1998]** Papadimitriou C., Raghavan P., Tamaki H., Vempala S. Latent Semantic Indexing: A probabilistic analysis // Proc. of ACM PODS. – Doi:10.1145/275487.275505. – ISBN 978-0897919968. S2CID 1479546.
- [pyLDavis, 2022]** Документация pyLDavis [Электронный ресурс]. – URL: <https://pyldavis.readthedocs.io/en/latest/index.html> (дата обращения: 18.09.2022).
- [Rosen-Zvi, 2004]** Rosen-Zvi M. The author-topic model for authors and documents // Proc. of the 20th Conference on Uncertainty in Artificial Intelligence, USA. 2004.
- [spaCy, 2022]** Industrial-Strength Natural Language Processing [Электронный ресурс]. – URL: <https://spacy.io/> (Дата обращения 17.09.2022).
- [STMT, 2022]** Stanford Topic Modeling Toolbox [Электронный ресурс]. – URL: <https://downloads.cs.stanford.edu/nlp/software/tmt/tmt-0.4/> (дата обращения: 17.09.2022).
- [Yuan et al., 2021]** Yuan J., He Z., Ming L., Lan D., Wray B. Neural Attention-Aware Hierarchical Topic Model [Электронный ресурс]. arXiv. Служба распространения и архив с открытым доступом. – URL: <https://arxiv.org/pdf/2110.07161.pdf> (дата обращения: 04.09.2022).
- [Zhijun, 2011]** Zhijun Y. Geographical topic discovery and comparison // Proc. of the 20th International Conference on World Wide Web. – Doi: 10.1145/1963405.1963443. – ISBN 9781450306324. S2CID 17883132.
- [Булатов, 2020]** Булатов В.Г. Методы оценивания качества и многокритериальной оптимизации тематических моделей в библиотеке TopicNet // спец. 05.13.18 Математическое моделирование, численные методы и комплексы программ, Москва.
- [Воронцов и др., 2012]** Воронцов К.В., Потапенко А.А. Регуляризация, робастность и разреженность вероятностных тематических моделей [Электронный ресурс]. Компьютерные исследования и моделирование // Электрон. журнал. – 2012. – № 4. – URL: <http://crm.ics.org.ru/> (дата обращения: 04.09.2022).

- [Гринева и др., 2009] Гринева М., Гринев М. Анализ текстовых документов для извлечения тематически сгруппированных ключевых терминов // Труды Института системного программирования РАН. – 2009. – Т. 16.
- [Коршунов и др., 2012] Коршунов А., Гомзин А. Тематическое моделирование текстов на естественном языке [Электронный ресурс]. Научная электронная библиотека «КиберЛенинка». – URL: <https://cyberleninka.ru/article/n/tematicheskoe-modelirovanie-tekstov-na-estestvennom-yazyke> (дата обращения: 04.09.2022).
- [Котов и др., 2005] Котов Э.М., Целых А.Н. Методы тематического анализа текстовой информации [Электронный ресурс]. Научная электронная библиотека «КиберЛенинка». – URL: <https://cyberleninka.ru/article/n/metody-tematicheskogo-analiza-tekstovoy-informatsii> (дата обращения: 04.09.2022).
- [Михайлов и др., 2015] Михайлов Д.В., Козлов А.П., Емельянов Г.М. Выделение знаний и языковых форм их выражения на множестве тематических текстов: подход на основе меры TF-IDF // Компьютерная оптика. – 2015. – Т. 39, № 3.
- [Нугуманова и др., 2016] Нугуманова А.Б., Бессмертный И.А., Пещина П., Байбурин Е.М. // Обогащение модели bag of words семантическими связями для повышения качества классификации текстов предметной области // Программные продукты и системы. – 2016. – № 2.
- [Сафонова, 2016] Сафонова А.В. Алгоритм сингулярного разложения матриц // Радиотехнические и телекоммуникационные системы. – 2016. – № 1 (21).
- [Фаткулин, 2015] Фаткулин Б.Г. Использование лингвистически ориентированных модулей на языке PYTHON для обработки больших текстовых массивов на восточных языках в целях эффективного сбора и обработки данных по отраслям востоковедческой тематики (на примере NLTK) // Вестник Южно-Уральского государственного университета, серия «Лингвистика». – 2015. – Т. 12. № 1.
- [Федоренко и др., 2018] Федоренко В.И., Киреев В.С. Использование методов векторизации текстов на естественном языке для повышения качества контентных рекомендаций фильмов // Современные наукоемкие технологии. – 2018. – № 3.



УДК 004.912+004.852+004.622

## МЕТОД АВТОМАТИЧЕСКОГО ПОСТРОЕНИЯ ОБУЧАЮЩИХ КОЛЛЕКЦИЙ ДЛЯ ЗАДАЧИ АБСТРАКТИВНОГО АННОТИРОВАНИЯ НОВОСТНЫХ СТАТЕЙ\* \*\*

Б.В. Добров (*dobrov\_bv@mail.ru*)  
Д.И. Чернышев (*chdanorbis@yandex.ru*)  
МГУ им. М.В. Ломоносова, Москва

Создание коллекции примеров для обучения систем абстрактного аннотирования является затратным процессом из-за больших временных затрат и высоких требований к квалификации экспертов необходимых для написания качественной аннотаций. Для задачи абстрактного аннотирования новостных кластеров предлагается новый метод создания коллекций для обучения нейросетевых методов аннотирования – ClusterVote, предназначенный моделировать особенности задачи путем учета информации в связанных документах. Метод может быть использован для формирования абстракттивных аннотаций различного уровня детализации, а также для получения экстрактивных аннотаций. Используя метод ClusterVote, была сформирована новая коллекция на английском и русском языках для обучения систем аннотирования новостных статей – Telegram News\*CV. Экспериментальные результаты показывают, что при определенных параметрах сформированные ClusterVote коллекции имеют схожие экстрактивные характеристики с такими известными наборами данных как CNN/Daily Mail и при этом обладают более высокими показателями “фактической достоверности” – воспроизведения в аннотациях именованных сущностей исходных текстов, а также их связей.

**Ключевые слова:** абстрактное аннотирование, новости, машинное обучение, нейросетевые технологии.

---

\* Работа выполнена при финансовой поддержке РФФ (проект № 21-71-30003).

\*\* Работа выполнена при финансовой поддержке Некоммерческого Фонда развития науки и образования “Интеллект”.

## Введение

Аннотирование текстов – одна из самых сложных и востребованных областей обработки естественного языка ([El-Kassas et al., 2021], [Ma et al., 2020]). Традиционно эту задачу представляют как сжатие текста без потери смысла и решают двумя подходам: экстрактивным (англ. – extractive summarization) и абстрактивным (англ. – abstractive summarization). При экстрактивном аннотировании извлекаются наиболее значимые фрагменты исходного текста. Абстрактивное аннотирование допускает возможность перефразирования исходного текста, что потенциально позволяет формировать аннотации в едином стиле, в случае обзорного (многодокументного) аннотирования – формировать связные аннотации, обобщающие тексты разных документов.

В настоящее время методы абстрактивного аннотирования в широких предметных областях базируются на нейросетевых подходах ([Gupta et al., 2019], [Kryściński et al., 2021]), которым для обучения требуются большие коллекции примеров с качественными аннотациями, в идеале формируемые экспертами в предметной области. Однако, объем работы и высокие требования к квалификации экспертов, необходимые для написания качественных аннотаций в широких предметных областях, делают такой подход очень затратным. Возникает задача автоматического формирования обучающих коллекций для аннотирования с заданными и контролируемыми свойствами.

Для решения этой задачи было предложено несколько универсальных автоматических подходов построения наборов данных для аннотирования. Общая идея таких подходов – отбор наиболее значимых предложений на основе статистических оценок и автоматическое формирование «псевдо-аннотаций», моделирующих аннотации, порождаемые экспертами. Такой подход показал себя эффективным для предобучения современных нейросетевых моделей абстрактивного аннотирования, таких как Pegasus [Zhang et al., 2020a] и PRIMERA [Xiao et al., 2022], которые показывают лучшие результаты в задаче даже при малом (меньше 1000) количестве обучающих примеров. Однако эвристические метрики, учитывающие только частотные зависимости, применяемые для оценки значимости предложений, не могут в полной мере отражать смысл исходных текстов, и поэтому псевдо-аннотации, полученные такими методами, могут значительно отличаться от сформированных экспертами. Для задачи аннотирования новостных кластеров (сообщений, посвященных описанию одного события – новостного сюжета) важно учитывать интегрированную информацию всех документов кластера.

Мы предлагаем новый метод для получения псевдо-аннотаций – ClusterVote. Идея метода ClusterVote состоит в том, что, выбирая наиболее цитируемые или упоминаемые предложения (с учетом возможного перефразирования) в кластере, мы можем получить экстрактивную аннотацию, которая бы полнее отражала усредненную точку зрения различных авторов текстов новостного сюжета.

В отличие от предыдущих подходов, наш метод способен создавать как экстрактивные, так и абстрактивные аннотации с различным уровнем детализации, учитывает структурные (структуру изложения информации) и жанрово-стилистические особенности текстов предметной области.

Используя этот подход, мы создали новый набор данных для аннотирования новостных статей Telegram News\*CV на основе данных соревнования Telegram Data Clustering Contest 2020<sup>1</sup>. Мы приводим оценку качества коллекций для различных вариантов метода и сравниваем их эффективность с ранее предложенным подходом псевдо-аннотирования для новостей [Chernyshev et al., 2021]. Эксперименты показывают, что при определенных значениях параметров метода ClusterVote, результирующие коллекции имеют схожие экстрактивные характеристики с такими коллекциями как CNN / Daily Mail [Nallapati et al., 2016], а обученные на этих коллекциях методы абстрактивного аннотирования обладают более высокой степенью “фактологической достоверности” (воспроизведения в аннотациях именованных сущностей исходных текстов, а также их связей) даже при достаточно низком уровне воспроизведения исходного текста.

## 1. Обзор литературы

Известно большое количество методов аннотирования [El-Kassas et al., 2021]. Абстрактивное аннотирование получило значительный прогресс с развитием методов нейросетевого глубокого обучения [Gupta et al., 2019].

### 1.1. Модели абстрактивного аннотирования

На данный момент лучшие результаты в задаче абстрактивного аннотирования показывают модели Pegasus [Zhang et al., 2020a] и PRIMERA [Xiao et al., 2022]. Качество моделей оценивается по двум основным показателям: точность воспроизведения эталонных образцов и “естественность” текста аннотации. Первое вычисляется на основе автоматических программных метрик вроде ROUGE [Lin, 2004], что отражает способность модели угадывать отдельные слова и фразы. Для применения модели на практике необходимо также чтобы текст аннотации был связным, соответствовал контексту исходной статьи и не противоречил известным фак-

---

<sup>1</sup> <https://contest.com/data-clustering-2>.

там и здравому смыслу. Ранние модели абстрактивного аннотирования часто сталкивались с проблемами подмены сущностей (например, “Новая Зеландия” может заменяться на “Ирландию”) [Cao et al., 2017] и “заикливания” процесса генерации [See et al., 2017], что приводило к повтору одного и того же предложения на протяжении всей аннотации. С появлением предобученных нейросетевых языковых моделей [Devlin et al., 2019], имеющим возможность учета глобального контекста, характеристика “естественности” свелась к непротиворечивости [Maynez et al., 2020], для оценки которой применяют различные метрики достоверности [Pagnoni et al., 2021].

Нейросетевые модели допускают эволюционное развитие, когда новые модели учитывают достижения предыдущих, добавляя улучшения в архитектуру нейронной сети или в методы обучения. Модель Pegasus является улучшением модели BART [Lewis et al., 2020], которая в свою очередь является модификацией архитектуры BERT [Devlin et al., 2019] для задачи генерации текста. BART, как и BERT, предобучен для языковых задач типа Cloze test (заполнение при обучении пропусков в тексте наиболее подходящими по контексту словами). Однако, если BERT обучен предсказывать не более одного слова в пропуске на основе построенного векторного представления входного текста, то BART рекурсивно декодирует это представление для восстановления пропущенных последовательностей произвольной длины. Благодаря такой особенности BART подходит для любых задач генерации текста, в том числе и аннотирования: на вход подается текст, где единственный пропуск обозначен сразу после текста, а модель обучается заполнять этот пропуск, сопоставляя свое предсказание с эталонной аннотацией.

Корень проблемы применения BART для задачи абстрактивного аннотирования заключается в постановке задачи восстановления предложений, на которой предобучается модель. В постановке решаемой BART задачи, модель обучается восстанавливать случайные предложения, а поскольку случайное предложение может быть уникальным по содержанию и не иметь связи с остальным текстом, это приводит к тому, что модель учится приоритизировать глобальные контекстные знания и игнорировать локальный контекст. Для задачи сжатия текста, в том числе и аннотирования, сохранение локального контекста часто важнее глобальных знаний, поэтому модель Pegasus использует специальный алгоритм отбора предложений для восстановления, который гарантирует наличие необходимой для восстановления информации в неотобранных предложениях. Благодаря такой постановке задачи предобучения модель Pegasus без учителя показывает 75% эффективности полностью обученной BART и достигает 90% менее чем за 100 примеров [Zhang et al., 2020a].

PRIMERA можно рассматривать как модификацию Pegasus для случая обзорного (многодокументного) аннотирования. В PRIMERA используется модификация BERT с разреженным механизмом внимания Longformer, которая позволяет обрабатывать в 4 раза больше входных слов. Также, для учета иерархии документов в задаче предобучения, предложения для восстановления в документе отбираются в соответствии с иерархией оценки значимости именованных сущностей, а в качестве входной информации используются тексты других документов. Это позволило модели PRIMERA улучшить результаты Pegasus на 15% в условиях обработки нескольких документов [Xiao et al., 2022].

В настоящее время, для русского языка пока неизвестно подобных PRIMERA и Pegasus специализированных моделей. Поэтому в настоящей работе для текстов на русском языке мы рассматривали ближайшие аналоги: мультязычную версию модели mBART и модель ruT5-large, которая является адаптацией модели T5 [Raffel et al., 2019] для русского языка. Модель T5 имеет ту же архитектуру, что и BART, но предобучена также на задачах восстановления порядка слов.

## **1.2. Наборы данных для аннотирования новостных статей**

Исторически первым набором данных для абстрактивного аннотирования была коллекция New York Times Annotated Corpus [Sandhaus, 2008] (более 600 тыс. пар статья-аннотация). Однако из-за ограниченного доступа, в научном сообществе для тестирования методов аннотирования большее распространение получила коллекция CNN/Daily Mail [Nallapati et al., 2016] (300 тыс. пар). Но эталонные аннотации набора CNN/Daily Mail имеют высокие показатели экстрактивности (часто повторяют фрагменты исходных текстов), поэтому в 2018 году был предложен набор Xsum [Narayan et al., 2018] (226 тыс. пар), который содержит аннотации длиной в одно предложение с новостного сайта BBC News. В том же году был предложен набор Newsroom [Grusky et al., 2018] (1.3 млн пар), который был автоматически собран путем выделения формальных структурных фрагментов в коде интернет-страниц различных новостных изданий. Следуя этому подходу, в 2020 году был опубликован первый набор для аннотирования русскоязычных новостных статей Gazeta [Gusev, 2020] (64 тыс. пар). В 2020 и 2021 году были опубликованы мультязычные наборы для аннотирования новостей MLSum [Scialo et al., 2020] и XLSum [Hasan et al., 2021], которые также содержат статьи на русском языке. На сегодняшний день суммарное количество русскоязычных пар статья-аннотация среди всех опубликованных наборов не превышает 180 тысяч.

Главный недостаток существующих автоматически собранных наборов данных – невысокое качество эталонных аннотаций. Было показано [Tejaswin et al., 2021], что наборы данных, собранные на основе формаль-

ной структуры веб-страниц, такие как Xsum и Newsroom содержат фрагменты, несвязанные с исходным текстом, или противоречивые аннотации. Несмотря на дополнительные усилия по фильтрации, коллекция Gazette также содержит достаточное количество некорректных эталонных аннотаций [Chernyshev et al., 2022].

### **1.3. Использование псевдо-аннотаций для машинного обучения методов аннотирования**

Наборы данных для обучения абстрактивного аннотирования различаются в первую очередь жанрово-стилистическими и структурными особенностями аннотаций. Поэтому для достижения максимальной эффективности систем аннотирования желательно формировать наборы данных под конкретные области применения. Однако привлечение экспертов для формирования новых коллекций слишком затратно. В качестве разумной альтернативы были разработаны автоматические методы построения обучающих коллекций на основе псевдо-аннотаций – аннотаций полученных автоматически на основе эвристического отбора наиболее значимых предложений текста.

Впервые эта идея была применена для обучения нейросетевой модели абстрактивного аннотирования TED [Yang et al., 2020], в которой, следуя принципу «перевернутая пирамида<sup>2</sup>», в соответствии с которым основное содержание новости излагается в начале текста, в качестве псевдо-аннотаций для новостных статей использовались первый абзацы. Но в работе [Zhang et al., 2020b] было показано, что такая стратегия менее эффективна для обучения, чем случайный выбор подмножества предложений, и взамен был предложен метод отбора предложений для псевдо-аннотации на основе их униграммной схожести с остальным текстом. Этот подход был позже дополнен иерархией оценки значимости именованных сущностей [Xiao et al., 2022], для моделирования информационной иерархии текста.

## **2. Формирование коллекции псевдо-аннотаций методом ClusterVote**

В качестве основы для экспериментов мы выбрали данные, предоставленные в рамках научно-технологического соревнования Telegram Data Clustering Contest 2020, которое проводила компания Telegram FZ-LLC. В рамках соревнования организатор предоставил коллекции из сотен тысяч новостей за каждый из нескольких дней на разных языках. Целью соревнования была кластеризация новостей по различным признакам: язык,

---

<sup>2</sup> [https://en.wikipedia.org/wiki/Inverted\\_pyramid\\_\(journalism\)](https://en.wikipedia.org/wiki/Inverted_pyramid_(journalism)).

категория, тема. Конечным результатом этого процесса должны были быть кластеры новостей, которые содержат только близкие по контексту статьи.

Используя предположение о тематической близости документов в новостном кластере, мы разработали метод ClusterVote, который ранжирует предложения в соответствии с популярностью представленной информации.

## 2.1. Набор данных Telegram Data Clustering Contest 2020

В наборе данных Telegram Data Clustering Contest представлены статьи на различных языках, опубликованные в течение 52 дней (ноябрь 2019, май 2020). Для исследования мы использовали только англоязычную (560 000 статей от 1346 издателей) и русскоязычную (600 000 статей от 1477 издателей) части. Telegram не предоставил разметку для данных, поэтому кластеры документов были собраны самостоятельно, используя решения победителей соревнования<sup>3</sup>. Для повышения качества кластеризации предварительно были удалены статьи длиной менее 50 слов.

Для оптимизации процесса кластеризации весь набор данных был разбит на 72-часовые сегменты с 24-часовым пересечением для учета возможной 48-часовой задержки, свойственной возможности появления обзорных новостных статей. Каждый сегмент был кластеризован алгоритмом DBSCAN [Ester et al., 1996] с косинусным расстоянием. Сначала все статьи в сегменте были кластеризованы в соответствии с именованными сущностями, а затем полученные кластеры были разбиты на подкластеры по tf-idf векторам, построенным по первым 4 предложениям, что моделировало структуру “перевернутой пирамиды”.

## 2.2. Метод ClusterVote

Разработан метод ClusterVote, в рамках которого предполагается, что текстовая близость новостных документов должна отражать близость представленных в новостях наборов фактов. Под “фактом” в настоящей работе понимается элементарная тройка <субъект-отношение-объект>, полученная на основе синтаксических правил, смоделированных признанными в научном сообществе инструментами такими как spaCy<sup>4</sup> и OpenIE<sup>5</sup>. Следуя этой логике, предполагается, что кластеры документов образуются вокруг некоторого ядра множества фактов, которое постоянно для всех документов в кластере. Расстояние конкретного факта до этого

---

<sup>3</sup> <https://github.com/IlyaGusev/tgcontest>.

<sup>4</sup> <https://github.com/explosion/spaCy>.

<sup>5</sup> <https://stanfordnlp.github.io/CoreNLP/openie.html>.

ядра отражает насколько заинтересовано сообщество авторов и читателей в нем. Самые близкие к ядру факты (в составе ядра) были оценены большинством авторов/издателей как значимые и поэтому являются основой объективной аннотации, в то время как наиболее отдаленные были оценены авторами как несущественные для читателя и поэтому опущены во многих документах.

Учитывая, что описание фактов часто отражает основной состав предложений текста, ядро фактов кластера может быть представлено как наиболее частый набор схожих предложений. Такое множество предложений можно определить, построив отображение предложений документа в предложения других документов кластера и последующим подсчетом поддержки – количества документов, содержащих предложения, близкие по упоминаемым фактам:

$$vote(s_i) = |\{D_k | s_i \notin D_k, \exists s_j \in D_k: s_j \equiv s_i\}|, \quad (3.1)$$

где  $D_k$  – документ из кластера,  $s_k$  – предложение некоторого документа.

Выбирая предложения, которые получили поддержку выше некоторого порога, можно получить аннотацию кластера с точки зрения конкретного документа. Изменение этого порога будет приводить к различной детализации аннотации. Для документа источника такая аннотация будет экстрактивной, а для других документов кластера абстрактивной. В некотором смысле можно считать, что документы кластера “голосуют” за предложения в других документах, что и определяет название метода ClusterVote.

Для группировки предложений использовался алгоритм DBSCAN с косинусным расстоянием и векторными представлениями, полученными с помощью SentenceTransformers из моделей paraphrase-mpnet-base-v2<sup>6</sup> и sbert\_large\_mt\_nlu\_ru<sup>7</sup> для английского и русского языков, соответственно. Поскольку содержание и число предложений в кластере не постоянно, порог слияния  $\epsilon$  устанавливается динамически с помощью метода локального перебора (grid search) для гарантии, что максимальное расстояние в кластере не будет превышать некоторый порог контекстной схожести  $d_{par}$ , который был подобран как 0.2 для английского и 0.3 для русского языков.

Для экспериментов были рассмотрены два характерных варианта метода:

- CV-full – порог поддержки  $t_{vote} = 1$ ;
- CV-max – порог поддержки  $t_{vote} = \max(\{vote(s_i) | s_i \in Article\})$ .

В каждой паре статей в качестве источника аннотации была выбрана статья с наименьшим числом общей поддержки.

<sup>6</sup> [https://www.sbert.net/docs/pretrained\\_models.html](https://www.sbert.net/docs/pretrained_models.html).

<sup>7</sup> [https://huggingface.co/sberbank-ai/sbert\\_large\\_mt\\_nlu\\_ru](https://huggingface.co/sberbank-ai/sbert_large_mt_nlu_ru).



Метод ClusterVote можно рассматривать как особый случай метода LexRank [Günes et al., 2004], который ранжирует предложения в соответствии с их центральностью в графе связей, но на невзвешенном графе с дополнительной фильтрацией ребер. Такая фильтрация представляет собой сведение графа близости предложений к графу кластеров, в котором единственные связанные компоненты – клики. Поскольку в таком графе центральность каждой вершины не зависит от вершин за пределами клик, то LexRank точно сохранит центральность внутри клик и таким образом получит то же ранжирование, что и ClusterVote.

### 2.3. Метрики качества

Для оценки качества наборов данных мы вычисляем традиционные экстрактивные метрики [Grusky et al., 2018]. Покрытие экстрактивными фрагментами – это доля слов исходного текста, которые сохранились в процессе аннотирования:

$$\text{Coverage}(A, S) = \frac{1}{|S|} \sum_{f \in F(A, S)} |f|, \quad (3.2)$$

где  $A$  – текст исходной статьи,  $S$  – текст аннотации и  $F(A, S)$  – это набор фрагментов аннотации, которые были напрямую извлечены из текста статьи. Эта метрика показывает лексическое сходство или количество синонимичных замен.

Плотность экстрактивных фрагментов – это средняя длина экстрактивного фрагмента  $f$  к которому принадлежит каждое слово в аннотации:

$$\text{Density}(A, S) = \frac{1}{|S|} \sum_{f \in F(A, S)} \sum_{w \in f} w \cdot |f| = \frac{1}{|S|} \sum_{f \in F(A, S)} |f|^2. \quad (3.3)$$

В отличие от покрытия плотность чувствительна к количеству экстрактивных фрагментов. Например, если у нас есть аннотация, состоящая из 60 слов, и общая длина экстрактивных фрагментов составляет 42 два слова, то плотность будет 29.4 в случае только одного фрагмента и 14.7 для двух фрагментов одинаковой длины, в то время как покрытие будет 0.7 вне зависимости от разбиения. Максимальное значение плотности равно длине аннотации, поэтому для снижения эффекта разницы длин в различных наборах данных метрику удобнее нормализовать. Нормализованная плотность показывает долю многословных парафраз, возникших в процессе аннотирования.

В работе [Maunez et al., 2020] было показано, что фактологические ошибки в обучающих примерах приводят к генерации недостоверных аннотаций при применении систем абстрактного аннотирования, поэтому

важно проверять их корректность. Для измерения достоверности текста не существует универсальной метрики, поскольку определение факта зависит от извлекаемого контекста. Для учета различных фактологических аспектов мы вычисляем несколько метрик.

Точность воспроизведения фраз отражает долю фраз, извлеченных из текста:

$$phrase_{acc}(A, S) = ROUGE-2_{prec}(S, A). \quad (3.4)$$

Высокий уровень экстрактивности фраз гарантирует, что предложения аннотации соответствуют содержанию исходного текста, в то время как низкий показывает чрезмерное перефразирование или наличие внешней информации.

Точность воспроизведения именованных сущностей – это доля именованных сущностей правильно воспроизведенных в аннотации:

$$NEO(A, S) = \frac{|NE(A) \cap NE(S)|}{|NE(S)|}, \quad (3.5)$$

где  $NE(T)$  – именованные сущности текста  $T$ . Именованные сущности определяют контекстное ядро текста и поэтому в большинстве случаев не допускают перефразирования. Точность представленных фактов – это доля троек <субъект-отношение-объект>, извлеченных из аннотации, которые поддерживаются исходным текстом:

$$fact_{acc}(A, S) = \frac{|Facts(A) \cap Facts(S)|}{|Facts(S)|}, \quad (3.6)$$

$fact_{acc}$  – это основной способ вычисления фактологической точности в аннотировании [Goodrich et al., 2019]. Для извлечения именованных сущностей и троек фактов для метрик мы использовали spaCy и OpenIE.

Основная проблема упомянутых метрик – они чувствительные к синонимичным заменам. Это делает их недостоверными в условиях частого перефразирования. Для учета таких случаев мы дополнительно измеряем точность по BERTScore [Zhang et al., 2020b], которая показала высокую корреляцию с мнением экспертов в сравнительном тестировании фактологических метрик FRANK<sup>8</sup> [Pagnoni et al., 2021].

---

<sup>8</sup> <https://frank-benchmark.herokuapp.com/>.

## 2.4. Статистика для англоязычной части коллекции

Применив метод ClusterVote к построенным ранее кластерам, мы получили 110 713 пар статья-аннотация для англоязычной части. Эту часть набора данных мы обозначили как Telegram News\*CV(EN)<sup>9</sup>.

Для оценки естественности псевдо-аннотаций мы проводим сравнение с существующими наборами с точки зрения метрик качества: CNN/Daily Mail и Newsroom. CNN/Daily Mail является стандартным набором для обучения и тестирования систем аннотирования, а Newsroom является ближайшим аналогом нашему методу, поскольку использует автоматический подход извлечения аннотации, выделяя маркеры эталонных аннотаций из HTML кода интернет-страницы. В дополнение мы также сравниваем аннотации полученные ClusterVote с ранее предложенным способом [Chernyshev et al., 2021] построения абстрактных псевдо-аннотаций на основе первого абзаца из статьи-аналога (обозначен как Pseudo Lead).

Примеры сравниваемых стратегий построения псевдо-аннотаций приведены в табл. 1. Выделенные в аннотации CV-full фрагменты пересекаются с CV-max, а подчеркнутые – с Pseudo Lead.

В табл. 2 приводится статистика по экстрактивным характеристикам наборов данных. Как можно заметить, тексты статей в нашем наборе данных значительно короче по сравнению с коллекциями CNN/DM и Newsroom. Тем не менее, несмотря на высокие показатели экстрактивного покрытия (coverage), нормализованная экстрактивная плотность у всех вариантов псевдо-аннотаций ниже, чем у Newsroom, причем самое низкое значение наблюдается у варианта CV-full. Поскольку CV-max является подмножеством CV-full, то такой результат говорит о том, что предложения с меньшей поддержкой имеют большую долю перефразированные фрагментов.

Таблица 1

### Примеры различных стратегий построения псевдо-аннотаций для английского языка

<b>Источник аннотации:</b>	<a href="https://www.vanguardngr.com/2020/04/ronaldinho-says-arrest-and-confinement-has-been-hard/">https://www.vanguardngr.com/2020/04/ronaldinho-says-arrest-and-confinement-has-been-hard/</a>
<b>Pseudo Lead:</b>	Kindly Share This Story: <u>Former Brazilian football star Ronaldinho said on Monday that his arrest and subsequent house arrest in Paraguay for using a false passport was “a very hard blow”.</u> “I would never have imagined myself in such a situation,” said the former Barcelona and Paris Saint-Germain striker in an interview with the Paraguayan newspaper ABC.

<sup>9</sup> Код метода ClusterVote для формирования коллекции Telegram News\*CV будет выложен в репозитории <https://github.com/dciresearch/ClusterVote>.

**CV-full:** Former Brazilian football star Ronaldinho said on Monday that his arrest and subsequent house arrest in Paraguay for using a false passport was “a very hard blow”. Ronaldinho and his brother Roberto de Assis Moreira are accused of entering Paraguay in possession of false passports. “I was completely caught off guard when I found out that these passports were not valid,” Ronaldinho said in his first public statement since his arrest two days after arriving in Asuncion at the beginning of March. Since April 7 they have been under house arrest at the Palmaroga Hotel in the historic centre of the Paraguayan capital. Ronaldinho said he hopes to be released “as soon as possible” after cooperating with the Paraguayan police investigation. Ronaldinho and his brother face up to five years in prison if found guilty.

**CV-max:** Ronaldinho and his brother Roberto de Assis Moreira are accused of entering Paraguay in possession of false passports. Since April 7 they have been under house arrest at the Palmira Hotel in the historic centre of the Paraguayan capital. Ronaldinho and his brother face up to five years in prison if found guilty.

В табл. 3 указана статистика по фактологическим метрикам. Для всех вариантов псевдо-аннотаций экстрактивные фактологические метрики выше, чем для аннотаций из популярных наборов данных, однако вариант Pseudo Lead имеет самые низкие показатели достоверности и чаще всех допускает ошибки при воспроизведении именованных сущностей.

Таблица 2

### Сравнение наборов данных для английского языка

Набор данных	Статья # слов	Аннотация # слов	Coverage	Density	
				raw	norm.
CNN/DM	781	56	89%	3,87	0,07
Newsroom	659	27	83%	9,51	0,36
Telegram News*CV (EN)					
Pseudo Lead	438	88	87%	26,10	0,29
CV-full		237	91%	43,43	0,18
CV-max		95	92%	28,81	0,30

Низкие показатели  $fact_{acc}$  в табл. 3 для CNN/Daily Mail связаны с самыми низкими экстрактивными характеристиками, поскольку метрика считает за ошибку несоответствие любой из компонент тройки факта <субъект-отношение-объект>. Поэтому и необходима метрика BERTScore. С точки зрения этой метрики CNN/Daily Mail имеет тот же уровень достоверности, что и Newsroom, однако, среди псевдо-аннотаций самым достоверным вариантом становится CV-full. Учитывая более низкие показатели экстрактивности по сравнению с CV-max, этот факт вновь указывает на наличие значительного перефразирования в предложениях с более низкой поддержкой.

Таблица 3

### Фактологические метрики для англоязычных наборов данных

Набор данных	<i>phrase<sub>acc</sub></i>	NEO	<i>fact<sub>acc</sub></i>	BERTScore
CNN/Daily Mail	49,78%	78,12%	9,39%	36,28%
Newsroom	53,89%	76,42%	38,39%	38,42%
Telegram News*CV (EN)				
Pseudo Lead	72,10%	75,53%	44,57%	63,41%
CV-Full	79,28%	80,18%	54,57%	<b>69,77%</b>
CV-Max	<b>83,09%</b>	<b>84,38%</b>	<b>61,90%</b>	63,15%

### 2.5. Статистика для русскоязычной части коллекции

Для русского языка – коллекция Telegram News\*CV(RU) – получилось 173 482 пар статья-аннотация. Для сравнения мы взяли набор Gazeta, поскольку он является единственным качественным набором для русского языка. Также, как и для английского мы приводим сравнение со стратегией псевдо-аннотирования Pseudo Lead. Примеры приведены в табл. 4.

Таблица 4

### Примеры различных стратегий построения псевдо-аннотаций для русского языка

<b>Источник аннотации:</b>	<a href="https://biwork.ru/news/koronavirus-v-altajskom-krae-dannye-na-2000-6-maa">https://biwork.ru/news/koronavirus-v-altajskom-krae-dannye-na-2000-6-maa</a>
<b>Pseudo Lead:</b>	<u>Информация по Алтайскому краю готовится на основе сводок Оперштаба края.</u> На два района расширилась география коронавируса в Алтайском крае. За минувшие сутки в Алтайском крае зарегистрировано 33 новых случая заболевания коронавирусной инфекцией.
<b>CV-full:</b>	<u>Информация по Алтайскому краю готовится на основе сводок Оперштаба края.</u> За минувшие сутки в Алтайском крае зарегистрировано 33 новых случая заболевания коронавирусной инфекцией. Среди новых случаев заболевания: жители Барнаула – 14 человек, Змеиногорского района – 9, Рубцовска – 2, Бийска – 1, Немецкого национального района – 1, Тальменского района – 1. Впервые зарегистрировано заболевание у трех жителей Ребрихинского и двух жителей Благовещенского районов. Из общего числа заболевших 56 случаев – завозных, в том числе 2 – из-за рубежа 430 случаев – по контакту с заболевшими. 84 – выявлены у больных внебольничной пневмонией, 94 – у заболевших ОРВИ, в 410 случаях – бессимптомное течение. Общее количество заболевших в регионе составляет 588 человек, 110 из них выздоровели и выписаны из клиник края, в том числе 2 пациента - за прошедшие сутки.
<b>CV-max:</b>	За минувшие сутки в Алтайском крае зарегистрировано 33 новых случая заболевания коронавирусной инфекцией. Общее количество заболевших в регионе составляет 588 человек, 110 из них выздоровели и выписаны из клиник края, в том числе 2 пациента - за прошедшие сутки.

Поскольку рассматриваемые метрики были разработаны без учета флективности русского языка, при вычислении метрик использовались лемматизированные версии текстов.

Статистика по наборам представлена в табл. 5. Можно заметить, что по своим характеристикам набор Gazeta схож больше с CNN / Daily Mail чем с Newsgroom, однако степень экстрактивности ещё ниже (нормализованная плотность экстрактивных фрагментов всего 0,05). Если сравнивать псевдо-аннотации, то они также более абстрактные, причем покрытие экстрактивными фрагментами даже ниже, чем для Gazeta. Псевдо-аннотации CV-max самые экстрактивные в данном случае, но разница с остальными вариантами пренебрежительно мала.

Возникает вопрос, на сколько будут эффективные фактологические метрики, если все варианты имеют еще меньшую степень экстрактивности чем аннотации из набора CNN / Daily Mail.

В табл. 6 представлены значения. Метрика  $fact_{acc}$  не может быть посчитана для русского языка в силу отсутствия стандартного (общедоступного) инструмента для извлечения троек фактов и поэтому отсутствует в таблице. Значение метрики  $phrase_{acc}$  значительно ниже в силу свободного порядка слов в русском языке, который допускает перестановки слов, не сохраняющие биграмы. С точки зрения NEO характеристики русскоязычных наборов эквивалентны англоязычным аналогам, но в данном случае CV-full оказывается наименее достоверным, находясь на том же уровне, что и Pseudo Lead. При этом BERTScore для всех выше, для Gazeta он составляет 75%, а для псевдо-аннотаций в районе 80%, с максимальным значением 81% также для CV-full.

Таблица 5

### Сравнение наборов данных для русского языка

Набор данных	Статья # слов	Аннотация # слов	Coverage	Density	
				raw	norm.
Gazeta	750	53	83%	2,46	0,05
Telegram News*CV (RU)					
Pseudo Lead	215	56	76%	6,77	0,12
CV-full		107	78%	11,19	0,10
CV-max		41	80%	7,41	0,18

Таблица 6

**Фактологические метрики для русскоязычных наборов данных**

Набор данных	<i>phrase<sub>acc</sub></i>	NEO	BERTScore
Gazeta	15,69%	78,40%	75,41%
Telegram News*CV (RU)			
Pseudo Lead	23,84%	76,15%	79,02%
CV-Full	<b>31,54%</b>	75,88%	<b>81,03%</b>
CV-Max	21,27%	<b>83,21%</b>	80,14%

**3. Оценка качества псевдо-аннотаций**

Пригодность использования набора данных для задачи аннотирования мы будем оценивать путем проведения экспериментов с существующими системами в традиционной постановке, когда часть коллекции используется для обучения, а оставшаяся часть для тестирования. При этом, если обученный на коллекции метод воспроизводит именованные сущности и фактологические взаимосвязи между ними, то такая коллекция признается приемлемой.

Более подробно – низкие показатели эффективности методов свидетельствует о том, что набор данных содержит некорректные примеры [Kang et al., 2020], которые смещают статистические характеристики задачи, делая невозможным нахождение оптимального решения, в то время как высокие говорят о тривиальности стратегии решения задачи [Narayan et al., 2018], из чего следует, что обученные на таком наборе данных системы не будут способны адаптироваться к неидеальным условиям применения на практике. При этом по мере улучшения основных показателей качества аннотирования во время обучения показатели фактологической достоверности должны не уменьшаться. В противном случае эталонные аннотации не удовлетворяют главному ограничению задачи – сохранению целостности ключевой информации [Maynez et al., 2020].

**3.1. Базовые модели экстрактивного аннотирования**

Во многих работах по аннотированию приводится сравнение с тремя базовыми методами: Lead-k, Oracle и TextRank.

Lead-k представляет собой простое извлечение первых  $k$  предложений исходного текста, что моделирует схему “перевернутая пирамида”.

Oracle – является способом оценки верхней границы эффективности. Идея заключается в последовательном отборе предложений исходного текста жадным алгоритмом, который отбирает предложения до тех пор, пока растёт  $n$ -граммная схожесть с эталонной аннотацией.

TextRank [Mihalcea et al., 2004] один из первых способов построения экстрактивной аннотаций, использующий тот же алгоритм что и LexRank, но предназначенный для аннотирования одного документа.

### 3.2. Параметры экспериментов

В модели Oracle максимизируется сумма ROUGE-1 и ROUGE-2. В модели Lead- $k$  параметр  $k = 3$  для Pseudo Lead и CV-max, и  $k = 6$  для CV-full. В методе TextRank размер аннотации равен среднему размеру аннотаций в наборе данных. Для чистоты эксперимента все нейросетевые модели абстрактивного аннотирования получают на вход только первые 1024 токена исходного текста статьи, что является пределом для всех моделей, кроме PRIMERA. Для генерации аннотаций используется алгоритм beam search с размером луча равным 5 и блокировкой триграмм. В качестве метрик качества измеряются ROUGE F1 и фактологические метрики *phrase<sub>acc</sub>*, NEO, *fact<sub>acc</sub>*, и BERTScore. Для русского языка метрики рассчитывались на лемматизированных версиях текстов.

### 3.3. Результаты для Telegram News\*CV(EN)

В табл. 7 приведены результаты для стратегии построения псевдо-аннотаций Pseudo-Lead. Как и ожидалось, для этого случая простое извлечение первых 3 предложений оказывается достаточным оптимальным решением. Однако разница с Oracle составляет всего 6%, что указывает на тривиальность такого варианта псевдо-аннотации. Поскольку настоящая аннотация обычно отличается от первых 3 предложений, то предобученные модели уступают тривиальной схеме даже после полного обучения.

В табл. 8 представлены результаты для стратегии CV-full. Несмотря на более длинные аннотации, верхняя граница эффективности систем аннотирования выше, чем в случае Pseudo Lead. Тривиальная стратегия Lead-6 уже не такая эффективная, а обученные абстрактные модели значительно превосходят её, хотя и существенно отстают от метода Oracle.

В табл. 9 отображены результаты для стратегии CV-max. Поскольку эта стратегия использует подмножество предложений CV-full, то результаты сопоставимы. Эффективность методов без учителя гораздо ниже, Lead-3 по ROUGE-2 (покрытие биграмм/фраз) всего 38%, а абстрактные модели без обучения показывают не более 24%. При этом показатели Oracle для обоих вариантов ClusterVote по этой же метрике на 10% выше, чем для Pseudo Lead, что говорит о том, что аннотации представляют собой комбинацию фрагментов исходного текста.



Таблица 7

**Результаты для стратегии Pseudo Lead на Telegram News\*CV(EN)**

Модель	ROUGE-1	ROUGE-2	ROUGE-L	# слов
<b>Экстрактивные</b>				
Lead-3	74,41%	64,31%	68,57%	84
Oracle	80,16%	70,67%	74,40%	85
TextRank	47,46%	30,89%	37,81%	92
<b>Абстрактивные (Без обучения)</b>				
PRIMERA	36,49%	19,33%	25,63%	146
Pegasus	42,68%	28,23%	34,62%	79
<b>Абстрактивные (Полное обучение)</b>				
PRIMERA	72,13%	62,66%	66,43%	102
Pegasus	76,17%	66,53%	70,68%	90

Таблица 8

**Результаты для стратегии CV-full на Telegram News\*CV(EN)**

Модель	ROUGE-1	ROUGE-2	ROUGE-L	# слов
<b>Экстрактивные</b>				
Lead-6	56,96%	47,49%	50,24%	176
Oracle	86,15%	78,70%	80,87%	227
TextRank	58,28%	44,60%	47,77%	218
<b>Абстрактивные (Без обучения)</b>				
PRIMERA	45,60%	30,25%	32,81%	146
Pegasus	38,00%	28,01%	31,95%	79
<b>Абстрактивные (Полное обучение)</b>				
PRIMERA	66,91%	57,24%	60,02%	162
Pegasus	65,47%	56,29%	58,80%	160

Таблица 9

**Результаты для стратегии CV-max на Telegram News\*CV(EN)**

Модель	ROUGE-1	ROUGE-2	ROUGE-L	# слов
<b>Экстрактивные</b>				
Lead-3	52,07%	38,60%	43,56%	84
Oracle	87,00%	81,14%	83,72%	85
TextRank	43,04%	26,73%	33,70%	92
<b>Абстрактивные (Без обучения)</b>				
PRIMERA	37,03%	21,60%	27,32%	146
Pegasus	39,12%	23,71%	30,87%	79
<b>Абстрактивные (Полное обучение)</b>				
PRIMERA	60,92%	51,36%	54,87%	114
Pegasus	64,91%	55,33%	59,09%	88

Распределения схем извлечения предложений методом Oracle изображено на рис. 1. Для Pseudo Lead распределение смещено в сторону первых 3 предложений, что объясняет низкую разницу между Oracle и Lead-3 для этого типа аннотаций. Для CV-full, наоборот, распределение практически равномерное для первой половины документа с постепенным затуханием вероятности извлечения для последних предложений. CV-max будучи подмножеством CV-full лишь незначительно сдвигает распределение, но в сторону распределения набора CNN/Daily Mail. Поскольку CNN/Daily Mail содержит аннотации, написанные человеком, это говорит о том, что CV-max отражает общий замысел экспертов, сохраняя логику повествования перевернутой пирамиды.

*Рис. 1. Сравнение распределений относительных позиций предложений, извлеченных методом Oracle для англоязычной части*

Для оценки корректности псевдо-аннотаций мы сравниваем результаты полученные нейросетевыми моделям в табл. 10.

Таблица 10

### **Фактологические метрики для англоязычных систем аннотирования**

Модель	$phrase_{acc}$	NEO	$fact_{acc}$	BERTScore
Без обучения				
PRIMERA	86,74%	90,69%	76,47%	54,65%
Pegasus	98,13%	98,74%	93,42%	70,16%
Обучение – Pseudo Lead				
PRIMERA	87,97%	85,74%	64,83%	73,10%
Pegasus	93,74%	90,12%	74,82%	74,62%
Обучение – CV-full				
PRIMERA	91,61%	88,64%	69,76%	79,78%
Pegasus	95,99%	93,26%	80,35%	<b>84,58%</b>
Обучение – CV-max				
PRIMERA	96,02%	95,04%	82,51%	75,95%
Pegasus	<b>96,44%</b>	<b>95,49%</b>	<b>83,10%</b>	79,71%

Если примеры обучающей коллекции содержали существенные фактологические несоответствия, то для их воспроизведения модели должны были обучаться “угадывать” правильные варианты и тем самым нарушать достоверность аннотации.

Как можно видеть, фактологические метрики сгенерированных аннотаций для всех случаев выше, чем для псевдо-аннотаций, на которых модели обучались. Однако в случае Pseudo Lead наблюдается падение всех метрик после обучения для обеих моделей. Учитывая, что именованные сущности не должны перефразироваться в аннотации, получается, что Pseudo Lead создает недостоверные эталоны. При CV-full NEO тоже падает, но не так существенно и при этом рост BERTScore в этом случае гораздо сильнее, что говорит о большей контекстной уместности этих ошибок. Учитывая, что при CV-max наблюдается рост по всем метрикам, то скорее всего среди предложений с низкой поддержкой, которые были включены в CV-full, есть недостоверные.

#### 3.4. Результаты для Telegram News\*CV(RU)

В табл. 11 приведены результаты стратегии Pseudo Lead для русского языка. Как и для случая английского языка, Lead-3 является оптимальным решением и отличается от Oracle всего на 6% по ROUGE-1. Но при этом верхняя граница по Oracle составляет 41%, что говорит либо о высокой абстрактности аннотаций, либо о наличии в них несоответствий с текстом статьи.

Как можно видеть в табл. 12, для CV-full более высокие экстрактивные показатели. Верхний порог по ROUGE-1 составляет 50%, а разница с Lead-6 уже 14%. Модели абстрактного аннотирования уже значительно превосходят по качеству тривиальное решение.

Таблица 11

#### Результаты для стратегии Pseudo Lead на Telegram News\*CV(RU)

Модель	ROUGE-1	ROUGE-2	ROUGE-L	# слов
<b>Экстрактивные</b>				
Lead-3	34,42%	15,58%	33,32%	59
Oracle	41,23%	20,32%	39,74%	53
TextRank	22,57%	9,25%	21,71%	71
<b>Абстрактные (Полное обучение)</b>				
mBART	32,99%	15,24%	31,98%	51
ruT5	34,28%	16,18%	33,35%	56

В табл. 13 наблюдается подобное поведение моделей для CV-max, но экстрактивность псевдо-аннотаций ниже, чем у Pseudo Lead. Стоит отметить, что в этом случае наблюдается одинаковая эффективность нейросетевых моделей абстрактивного аннотирования, что говорит об устойчивости по данным оптимальной стратегии решения.

Таблица 12

**Результаты для стратегии CV-full на Telegram News\*CV(RU)**

Модель	ROUGE-1	ROUGE-2	ROUGE-L	# слов
<b>Экстрактивные</b>				
Lead-6	36,52%	18,18%	34,68%	116
Oracle	50,16%	29,86%	48,20%	98
TextRank	34,95%	17,21%	33,23%	124
<b>Абстрактивные (Полное обучение)</b>				
mBART	37,98%	15,98%	36,29%	102
ruT5	39,76%	21,50%	38,29%	109

Таблица 13

**Результаты для стратегии CV-max на Telegram News\*CV(RU)**

Модель	ROUGE-1	ROUGE-2	ROUGE-L	# слов
<b>Экстрактивные</b>				
Lead-3	21,13%	8,19%	20,43%	59
Oracle	36,77%	18,95%	35,90%	43
TextRank	17,57%	7,21%	17,15%	47
<b>Абстрактивные (Полное обучение)</b>				
mBART	26,74%	13,29%	26,36%	42
ruT5	26,96%	12,48%	26,38%	44

На рис. 2 изображено распределение стратегии Oracle. Несмотря на схожесть Gazeta с CNN/Daily Mail, оптимальная стратегия отказывается смещена в сторону первого предложения.

*Рис. 2. Сравнение распределений относительных позиций предложений, извлеченных методом Oracle для Telegram News\*CV(RU)*

Подобное распределение наблюдается для Pseudo Lead, однако вероятность извлечения предложений резко падает уже после первой четверти. Для CV-full и CV-max наблюдается мультимодальность, причем более строгие критерии отбора CV-max практически не изменяют распределение, что говорит о равнозначимости позиции в тексте при определении степени важности для русского языка.

В табл. 14 приведены фактологические метрики. Из-за низкой степени экстрактивности псевдо-аннотаций  $phrase_{acc}$  сохраняется низким после обучения.

Таблица 14

#### Фактологические метрики для русскоязычных систем аннотирования

Модель	$phrase_{acc}$	NEO	BERTScore
Обучение – Pseudo Lead			
mBART	37,07%	65,57%	81,97%
ruT5	31,27%	61,75%	80,16%
Обучение – CV-full			
mBART	36,86%	64,66%	79,60%
ruT5	<b>44,22%</b>	65,49%	<b>83,26%</b>
Обучение – CV-max			
mBART	28,30%	69,77%	81,47%
ruT5	28,19%	<b>72,12%</b>	81,50%

При этом показатели BERTScore высокие во всех случаях. Минимальный NEO наблюдается для моделей обученных на Pseudo Lead, а максимальный при обучении на CV-max несмотря на то, что  $phrase_{acc}$  на 10% ниже для последнего, хотя разница самих аннотаций по этой метрике в наборе данных была менее 2%. Учитывая, что достоверность аннотаций модели ruT5 обученной на Pseudo Lead существенно ниже, чем для mBART, а значения ROUGE выше, то можно утверждать, что псевдо-аннотации этого типа содержат неуместную информацию и непригодны для обучения моделей абстрактного аннотирования. В случае CV-full и CV-max не наблюдается такое противоречие в метриках, но рост NEO при падении  $phrase_{acc}$  в случае CV-max говорит о том, что парафразы и прямые цитаты одинаково оцениваются алгоритмом ClusterVote для русского языка, а также о большей поддержке фактологически правильных предложений.

#### Заключение

Предложен новый метод ClusterVote для автоматического построения обучающей коллекции для задачи аннотирования новостных документов, который в отличие от предыдущих подходов создает псевдо-аннотации с учетом связей с другими документами кластера. Используя этот метод, мы соз-

дали новую коллекцию Telegram News\*CV, которая включает аннотации для англоязычных и русскоязычных новостных статей. Мы протестировали два варианта ClusterVote: CV-full – предложения, которые получили поддержку хотя бы от одного документа кластера, CV-max – предложения с максимальной поддержкой в кластере. Мы сравнили эти подходы с нашим предыдущим методом Pseudo Lead, который использует особенности новостного стиля “перевернутая пирамида” и статьи-аналоги.

Согласно статистике на основе рассмотренных метрик, вариант CV-max формирует самые экстрактивные и фактически достоверные аннотации. Для английского языка этот тип псевдо-аннотации оказывается наиболее приближенным к коллекциям, полученным с привлечением экспертов, показывая то же распределение оптимальной экстрактивной стратегии, что и эталонный набор данных CNN/Daily Mail. Для русского языка это распределение сильнее смещено в сторону первых предложений даже для оригинальных аннотаций из набора Gazeta, но для псевдо-аннотаций метода ClusterVote этот эффект значительно меньше.

Результаты экспериментов показывают, что псевдо-аннотации ClusterVote, в отличие от псевдо-аннотаций Pseudo Lead, учат нейросетевые модели абстрактивного аннотирования сохранять фактологическую достоверность. При этом самые устойчивые и достоверные результаты дают модели, обученные на варианте CV-max. CV-full незначительно уступает по среднему качеству, но превосходит по показателям BERTScore, что говорит о пригодности для обучения моделей любых промежуточных вариантов детализации псевдо-аннотации ClusterVote.

## Список литературы

- [Cao et al., 2017] Cao Z., Wei F., Li W., and Li S. Faithful to the original: fact-aware neural abstractive summarization // Proceedings of the 32nd AAAI Conference on Artificial Intelligence and 30th Innovative Applications of Artificial Intelligence Conference and 8th AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'18/IAAI'18/EAAI'18). – 2018.
- [Chernyshev et al., 2021] Chernyshev D. and Dobrov B. Abstractive Summarization of Russian News Learning on Quality Media // Analysis of Images, Social Networks and Texts. AIST 2020. Lecture Notes in Computer Science. – 2021.
- [Chernyshev et al., 2022] Chernyshev D. and Dobrov B. Improving Neural Abstractive Summarization with Reliable Sentence Sampling // Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL 2021. Communications in Computer and Information Science. – 2022. – Vol. 1620.
- [Devlin et al., 2019] Devlin J., Chang M.-W., Lee K. and Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. – 2019. – Vol. 1.

- [El-Kassas et al., 2021] El-Kassas W.S., Salama C.R., Rafea A.A. and Mohamed H.K. Automatic text summarization: A comprehensive survey // *Expert Systems with Applications*. – 2021. – 165. – 113679.
- [Ester et al., 1996] Ester M., Kriegel H., Sander J. and Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise // *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96)*. – 1996.
- [Goodrich et al., 2019] Goodrich B., Rao V., Liu P.J. and Saleh M. Assessing The Factual Accuracy of Generated Text // *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. – 2019.
- [Grusky et al., 2018] Grusky M., Naaman M. and Artzi Y. Newsroom: A Dataset of 1.3 Million Summaries with Diverse Extractive Strategies // *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. – 2018. – Vol. 1. .
- [Gupta et al., 2019] Gupta S. and Gupta S.K. Abstractive summarization: An overview of the state of the art // *Expert Systems with Applications*. – 2019. – 121. – P. 49-65.
- [Günes et al., 2004] Günes E. and Radev D. R. LexRank: graph-based lexical centrality as salience in text summarization // *Journal of Artificial Intelligence Research*. – 2004.
- [Gusev, 2020] Gusev I., Dataset for Automatic Summarization of Russian News // *arXiv:2006.11063*. – 2020.
- [Hasan et al., 2021] Hasan T., Bhattacharjee A., Islam M. S. et al. XL-Sum: Large-Scale Multilingual Abstractive Summarization for 44 Languages // *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. – 2021.
- [Kang et al., 2020] Kang D. and Hashimoto T. Improved Natural Language Generation via Loss Truncation // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. – 2020.
- [Kryściński et al., 2021] Kryściński W., Rajani N. F., Agarwal D., et al. BookSum: A Collection of Datasets for Long-form Narrative Summarization // *arXiv:2105.08209*. – 2021.
- [Lewis et al., 2020] Lewis M., Liu Y., Goyal N. et al. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. – 2020.
- [Lin, 2004] Lin C.-Y. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out* // *Association for Computational Linguistics*. – 2004.
- [Ma et al., 2020] Ma C., Zhang W.E. Guo M. et al. Multi-document summarization via deep learning techniques: A survey. *ACM Computing Surveys (CSUR)*. – 2020.
- [Maynez et al., 2020] Maynez J., Narayan S., Bohnet B. and McDonald R. On Faithfulness and Factuality in Abstractive Summarization // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. – 2020.
- [Mihalcea et al., 2004] Mihalcea R. and Tarau P. TextRank: Bringing Order into Texts // *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. – 2004.
- [Nallapati et al., 2016] Nallapati R., Zhou B., Gulcehre C. and Xiang B. Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond // *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*. – 2016.

- [Narayan et al., 2018] Narayan S., Cohen S. B. and Lapata M. Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. – 2018.
- [Pagnoni et al., 2021] Pagnoni A., Balachandran V. and Tsvetkov Y. Understanding Factuality in Abstractive Summarization with FRANK: A Benchmark for Factuality Metrics // Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. – 2021.
- [Raffel et al., 2019] Raffel C., Shazeer N., Roberts A. et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer // Journal of Machine Learning Research. – 2019.
- [Sandhaus, 2008] Sandhaus E., The New York Times Annotated Corpus // Linguistic Data Consortium. – 2008.
- [Scialo et al., 2020] Scialo T., Dray P.-A., Lamprier S. et al. MLSUM: The Multilingual Summarization Corpus // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). – 2020.
- [See et al., 2017] See A., Liu P.J. and Manning C.D. Get To The Point: Summarization with Pointer-Generator Networks // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). – 2017.
- [Tejaswin et al., 2021] Tejaswin P., Naik D. and Liu P. How well do you know your summarization datasets? // Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. – 2021.
- [Xiao et al., 2022] Xiao W., Beltagy I., Carenini G. and Cohan A. PRIMERA: Pyramid-based Masked Sentence Pre-training for Multi-document Summarization // Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). – 2022.
- [Yang et al., 2020] Yang Z., Zhu C., Gmyr R. et al. TED: A Pretrained Unsupervised Summarization Model with Theme Modeling and Denoising // Findings of the Association for Computational Linguistics: EMNLP 2020. – 2020.
- [Zhang et al., 2020a] Zhang J., Zhao Y., Saleh M. and Liu P. PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization // Proceedings of the 37th International Conference on Machine Learning. – 2020.
- [Zhang et al., 2020b] Zhang T., Kishore V., Wu F. et al. BERTScore: Evaluating Text Generation with BERT // International Conference on Learning Representations. – 2020.



УДК 004.822:004.89

## ОПЫТ ИСПОЛЬЗОВАНИЯ ПАТТЕРНОВ СОДЕРЖАНИЯ ПРИ РАЗРАБОТКЕ ОНТОЛОГИЙ НАУЧНЫХ ПРЕДМЕТНЫХ ОБЛАСТЕЙ

Ю.А. Загорулько (*zagor@iis.nsk.su*)

Г.Б. Загорулько (*zagor@iis.nsk.su*)

Институт систем информатики им. А.П. Ершова СО РАН,  
Новосибирск

В настоящее время онтологии широко применяются для формализации и систематизации знаний и данных в различных предметных областях. Для облегчения и ускорения процесса разработки онтологий в последнее время используются паттерны онтологического проектирования (паттерны ОП), представляющие собой формальные описания проверенных на практике решений типовых проблем онтологического моделирования. Наиболее востребованным видом паттернов ОП являются паттерны содержания, поддерживающие единообразное и непротиворечивое представление всех сущностей моделируемой предметной области. В статье дается краткий обзор паттернов ОП, предлагается методология построения онтологий научных предметных областей (НПО), главной особенностью которой является использование базовых онтологий и паттернов содержания, а также анализируется опыт использования данной методологии при построении онтологий различных НПО.

**Ключевые слова:** научная предметная область, разработка онтологий, паттерны онтологического проектирования, паттерны содержания, специализация паттернов.

### Введение

В связи с тем, что онтологии показали себя удобным и эффективным средством формализации и систематизации знаний и данных в различных предметных областях [Кузнецов и др., 2010], существует насущная потребность в эффективных методах и инструментах построения онтологий. Особенно остро такая потребность ощущается при построении онтологий

научных предметных областей (НПО), к которым обычно относят предметные области, охватывающие некоторую научную дисциплину или область научных знаний во всех ее аспектах [Боргест, 2022].

Современные методы разработки онтологий [Sattar et al., 2020] можно разбить на три группы: (1) методы, обеспечивающие разработку онтологии “с нуля”, (2) методы создания онтологий из “готовых блоков”, (3) методы автоматического построения онтологий.

Методы разработки онтологии “с нуля” [Brusa et al., 2008; Цуканова, 2015; De Nicola et al., 2016; Кузнецов и др., 2016] являются самыми трудоемкими и сложными. Они требуют привлечения опытных специалистов в области онтологического инжиниринга.

Вторая группа включает менее трудоемкие методы [Gangemi et al., 2009; Suárez-Figueroa et al., 2015; Загорулько и др., 2016], использующие заранее созданные базовые онтологии и/или их фрагменты (паттерны), которые можно специализировать под конкретную предметную область. Достоинством этих методов является то, что они позволяют вовлечь в процесс построения и сопровождения онтологий экспертов в предметных областях, что является одним из важнейших современных требований к процессу разработки систем, основанных на знаниях [Грибова и др., 2022].

Методы автоматического построения онтологий на основе тематического корпуса текстов [Asim et al., 2018] или web-ресурсов [Singh et al., 2019] представляются наиболее перспективными и наименее трудозатратными, но при их использовании “в чистом виде” не удастся построить онтологии НПО хорошего качества [Ganino et al., 2018]. В связи с этим на практике такие методы часто используются для автоматического пополнения онтологий, построенных «ручными» методами [Ganino et al., 2018; Загорулько и др., 2021a].

В данной статье мы рассмотрим метод построения онтологий, ориентированный на инженеров знаний и экспертов предметной области и базирующийся на использовании паттернов онтологического проектирования (паттернов ОП) [Gangemi et al., 2009; Blomqvist et al., 2016; Zagorulko et al., 2018], которые представляют собой формальные описания проверенных на практике решений типичных проблем онтологического моделирования. В частности, в качестве паттернов могут выступать фрагменты базовых онтологий, представленные на языке OWL [Antoniou et al., 2009]. Таким образом, данный метод относится ко второй группе.

В статье дается краткий обзор паттернов онтологического проектирования, анализируются проблемы их применения, описывается предложенная авторами методология построения онтологий НПО, в которой используются базовые онтологии и паттерны ОП, а также опыт ее применения при разработке онтологий для различных НПО.

## **1. Паттерны онтологического проектирования и проблемы их применения**

Как было сказано выше, паттерны ОП служат для описания решений проблем, часто возникающих при разработке онтологий [Gangemi et al., 2009]. В зависимости от типа проблем, для решения которых предназначены паттерны онтологического проектирования, различают структурные паттерны, паттерны соответствия, паттерны содержания, паттерны логического вывода, паттерны представления и лексико-синтаксические паттерны. (Такая типология паттернов была предложена в рамках проекта NeOn [NeOn Project, 2022].)

При разработке онтологий инженерами знаний и экспертами в предметных областях в основном используются структурные паттерны, паттерны содержания и паттерны представления.

Структурные паттерны либо фиксируют способы решения проблем, вызванных ограничениями выразительных возможностей языков описания онтологий, либо задают общую (модульную) структуру и вид онтологии. Паттерны первого типа называются логическими паттернами, паттерны второго типа – архитектурными паттернами.

Паттерны содержания представляют собой описания типовых фрагментов онтологий, на основе которых могут строиться онтологии различных предметных областей.

Паттерны представления определяют рекомендации по именованию и аннотированию элементов онтологии, применение которых должно повысить читаемость онтологии, а также удобство и простоту ее использования.

На данный момент существует несколько каталогов паттернов ОП [ODPA, 2022; ODPS, 2009], доступных в сети Интернет. Наиболее представительный из них размещен на портале Ассоциации ODPA (Association for Ontology Design & Patterns) [ODPA, 2022], созданном в рамках проекта NeOn [NeOn Project, 2022].

Несмотря на то, что использование паттернов ОП позволяет экономить человеческие ресурсы и повысить качество разрабатываемых онтологий, они пока не нашли широкого практического применения из-за ряда проблем, возникающих при их использовании.

Первая проблема связана с трудностями повторного использования паттернов ОП, разработанных другими разработчиками, – зачастую разработчику новой онтологии трудно понять семантику, которую заложили в тот или иной паттерн его авторы.

Вторая проблема вызвана отсутствием удобных инструментов разработки онтологий, поддерживающих использование паттернов ОП. Здесь можно отметить плагины для инструмента разработки онтологий проекта

NeOn [Suárez-Figueroa et al., 2015] и редактора онтологий WebProtégé [Hammar, 2015]. Однако первый плагин доступен только для участников проекта NeOn, а второй может использоваться только в редакторе WebProtégé, который не очень популярен среди разработчиков онтологий из-за его ограниченной функциональности (по сравнению с настольной версией Protégé, для которой, кстати, не разработан подобный плагин).

Третья проблема связана с тем, что паттерны описываются и применяются отдельно и не составляют единой системы, что затрудняет их совместное использование.

Еще одна проблема состоит в том, что в широком доступе отсутствуют паттерны, которые можно было бы использовать при разработке онтологий научных предметных областей.

## **2. Методология разработки онтологий научных предметных областей**

Рассматриваемая в данной статье методология [Zagorulko et al., 2018] поддерживает построение онтологии НПО на основе базовой онтологии НПО, содержащей описания наиболее общих сущностей, характерных для большинства научных предметных областей. Так как онтология любой НПО содержит не только описания присущих ей системы понятий, решаемых задач и используемых методов, но и описания релевантных информационных ресурсов, базовую онтологию НПО содержательно можно разделить на четыре онтологии: (1) онтологию научного знания, (2) онтологию научной деятельности, (3) базовую онтологию задач и методов, (4) базовую онтологию информационных ресурсов.

Онтология научного знания содержит классы, задающие структуры для описания научных понятий, входящих в любую НПО. Такими понятиями являются Раздел науки, Объект исследования, Предмет исследования, Метод исследования, Научный результат, Задача, Алгоритм и др.

Онтология научной деятельности включает классы, относящиеся к организации научно-исследовательской деятельности, такие как Персона, Организация, Событие, Научная деятельность, Проект, Публикация и др.

Базовая онтология информационных ресурсов в качестве основного класса включает класс Информационный ресурс, набор атрибутов и связей которого основан на стандарте Dublin core [DCMI, 2022]. На основе данного класса могут быть описаны релевантные моделируемой НПО информационные ресурсы: базы данных, онтологии, коллекции документов, сайты организаций, персон и проектов.

Базовая онтология задач и методов включает классы и отношения, с помощью которых могут быть описаны задачи, решаемые в данной НПО, методы их решения и реализующие их программные компоненты и алгоритмы.

Все базовые онтологии имеют спецификации на языке OWL [Antoniou et al., 2009]. Для наиболее важных понятий базовой онтологии НПО разработаны паттерны содержания, также реализованные на языке OWL.

Так как данная методология поддерживает разработку онтологий инженерами знаний и экспертами, она предлагает систему паттернов ОП, которая включает три типа паттернов: паттерны представления, структурные паттерны и паттерны содержания.

Паттерны представления задают принятые в данной методологии правила именования и аннотирования элементов онтологии, подобные правилам, принятым в сообществе онтологического моделирования [Noy et al, 2001].

Структурные логические паттерны предназначены для представления сложных сущностей и конструкций, актуальных при построении онтологий НПО и для описания которых в языке OWL нет подходящих выразительных средств. В частности такие паттерны разработаны для представления областей допустимых значений аргументов, многоместных и атрибутированных отношений (бинарных отношений с атрибутами) [Загорулько и др., 2017].

Наиболее важную роль в рассматриваемой методологии играют паттерны содержания, которые используются для единообразного и непротиворечивого представления используемых в НПО понятий и их свойств. Как было сказано выше, такие паттерны разработаны для понятий, присутствующих в большинстве НПО. Благодаря этому, разработка онтологии конкретной НПО в основном сводится к специализации (настройке) паттернов содержания на эту область и построению на их основе фрагментов целевой онтологии.

Построение онтологии конкретной НПО с использованием базовых онтологий и системы паттернов ОП выполняется в два этапа:

1. Построение компонентов онтологии НПО на основе базовых онтологий путем их доработки и развития. На этом этапе выполняется специализация представленных в базовых онтологиях паттернов содержания и структурных логических паттернов на конкретную НПО.

2. Пополнение онтологии НПО путем конкретизации (означивания) паттернов содержания и структурных логических паттернов, представленных в базовых онтологиях или полученных из них путем их специализации на конкретную НПО.

При этом онтология области знаний строится на основе паттернов, представленных в базовых онтологиях научного знания и научной деятельности, онтология задач и методов – на основе паттернов базовой онтологии задач и методов, онтология научных интернет-ресурсов – на основе паттернов базовой онтологии информационных ресурсов.

Специализация паттерна ОП состоит в его переименовании, добавлении в него новых свойств (атрибутов и отношений) и/или в уточнении уже описанных в паттерне имен свойств и их областей значений.

Конкретизация паттерна ОП состоит в подстановке в него конкретных значений свойств и добавлении полученного фрагмента онтологии в создаваемую онтологию.

При разработке онтологии конкретной НПО в основном используются паттерны содержания, большая часть которых получена путем специализации паттернов содержания, образующих базовую онтологию НПО.

В качестве редактора онтологий и паттернов ОП на данный момент используется популярный редактор Protégé. Для поддержки использования паттернов ОП предоставляется редактор данных, позволяющий пополнять онтологию НПО путем конкретизации паттернов содержания.

### **3. Опыт использования методологии при разработке онтологий различных НПО**

Данная методология показала свою практическую полезность при разработке онтологий различных научных предметных областей: «Поддержка принятия решений» [Загорулько и др., 2021], «Решение вычислительно сложных задач математической физики на суперкомпьютерах» [Glinskiy et al., 2020], «Физика плазмы» [Snytnikov et al., 2020] и др. Рассмотрим процесс разработки онтологий для некоторых из них.

#### **3.1. Разработка онтологии предметной области «Решение вычислительно сложных задач математической физики на суперкомпьютерах»**

При разработке данной онтологии уточнялись и детализировались понятия базовой онтологии научного знания (*Объект исследования, Метод исследования, Раздел науки, Задача, Алгоритм*), а вместе с этим выполнялась их специализация на данную НПО. В результате в онтологию были введены такие сущности, как *Физический объект* и *Физическое явление*, *Фундаментальный закон природы*, *Физическая модель*, *Математическая модель*, *Система уравнений*, *Численный метод*, *Параллельный алгоритм* и другие. Для этих понятий были разработаны соответствующие паттерны содержания.

Рассмотрим в качестве примера паттерн для описания численных методов (рис. 1).

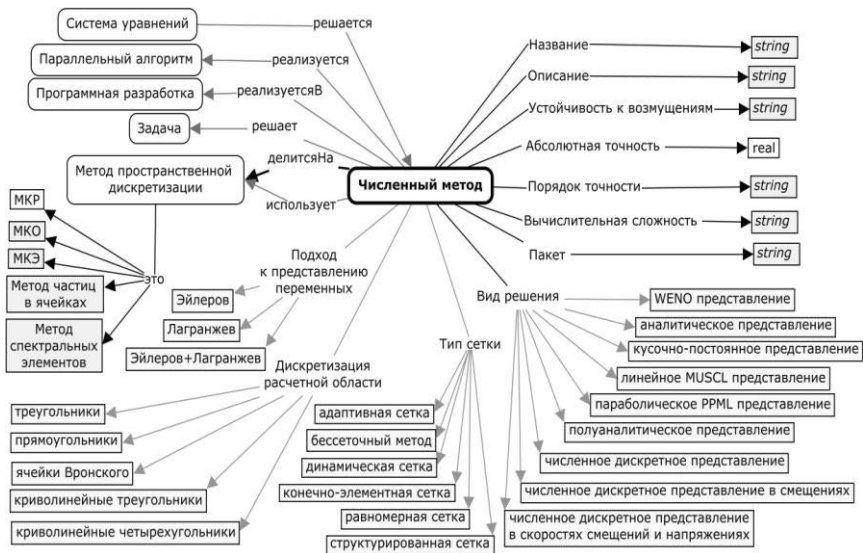


Рис. 1. Паттерн содержания «Численный метод»

С содержательной точки зрения такой паттерн представляет собой семантическую окрестность центрального понятия, которым в данном случае является класс *Численный метод*. Для этого понятия определены свойства, которые представляются в виде атрибутов или отношений. При этом атрибуты могут представляться либо как Data Properties – для тех свойств, значения которых имеют стандартный тип данных (*Название, Описание, Устойчивость к возмущениям, Абсолютная точность, Порядок точности, Вычислительная сложность*), либо как Object Properties – для свойств со значениями из перечислимого типа данных (*Вид решения, Тип сетки, Дискретизация расчетной области, Подход к представлению решения*). При этом сами перечислимые типы данных реализуются паттернами для представления областей допустимых значений.

Отношения задают связи объектов рассматриваемого класса с объектами других классов и представляются как Object Properties. В частности, в этом паттерне задаются связи с задачами, которые решает данный численный метод, с реализующим его параллельным алгоритмом, с решаемыми им системами уравнений и др.

Приведем также паттерн содержания для описания параллельных алгоритмов (рис. 2).



Рис. 2. Паттерн содержания для описания параллельных алгоритмов

В этом паттерне определены свойства параллельного алгоритма. Центральным понятием этого паттерна является класс *Параллельный алгоритм*. Он также имеет свойства, значения которых имеют стандартный тип данных (*Название*, *Описание*, *Оценка масштабируемости* и др.), и свойство *Декомпозиция расчетной области* со значениями из перечислимого типа данных. В этом паттерне параллельный алгоритм можно связать отношениями с целевой архитектурой, технологиями параллельного программирования, с используемой им структурой данных, с его кодом и т.п. При этом некоторые отношения могут иметь собственные атрибуты (*имеетВход*, *имеетВыход*, *определяется*).

Естественно, что для описания НПО в полном объеме предоставленных методологией паттернов может оказаться недостаточно. В связи с этим может потребоваться разработка паттернов содержания, специфичных для конкретной НПО. Такими паттернами для данной онтологии являются *Целевая архитектура*, *Технология параллельного программирования*, *Программная разработка* и др.

На рис. 3 представлен паттерн, описывающий понятие «Целевая архитектура».

На основе рассмотренной выше онтологии были разработаны онтологии нескольких областей математической физики, таких как астрофизика, геофизика и физика плазмы.

При описании этих областей использовались паттерны содержания, представленные в данной онтологии. На рис. 4 представлена конкретизация паттерна «Численный метод» для описания *Метода частиц в ячейках* при разработке онтологии физики плазмы [Snytnikov et al., 2020].





лей и задавать связи объектов онтологии с данными из внешних источников, которые могут использоваться при решении задач поддержки принятия решений (ППР). Наряду с понятиями, связанными с процессом принятия решений и методами его поддержки, предметная область «ППР» рассматривает и вопросы, связанные с программной реализацией методов, а также с осуществляемой в её рамках деятельностью.

Для специфичных для данной предметной области понятий (*Проблемная ситуация, Этап принятия решения* др.) были разработаны паттерны содержания «с нуля». Основная же часть паттернов содержания, использованных при разработке онтологии ППР, была получена путем специализации паттернов ОП, входящих в базовую онтологию НПО.

На рис. 5 показан паттерн для описания метода ППР, полученный путем специализации паттерна «Метод исследования» из базовой онтологии научного знания.

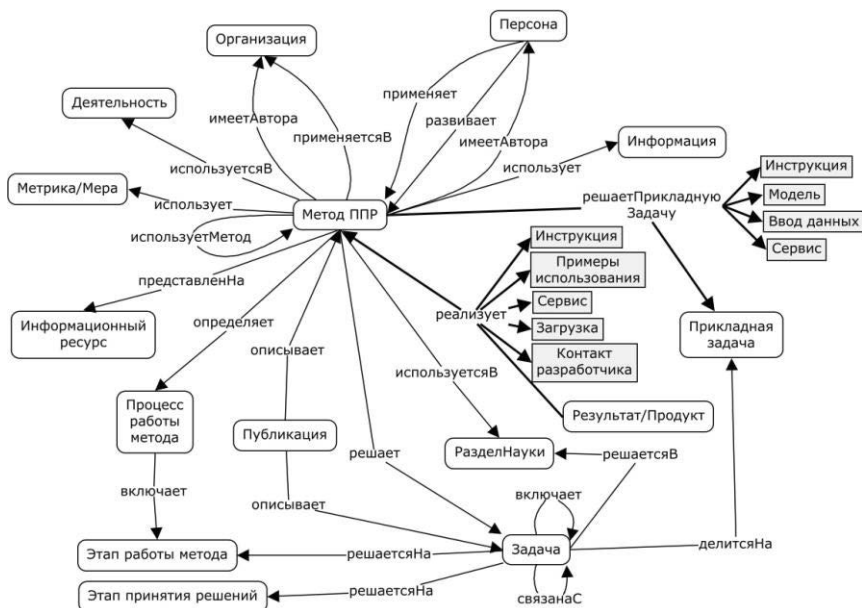


Рис. 5. Паттерн содержания «Метод ППР»

Паттерн содержания «Метод ППР» использовался для добавления в онтологию методов поддержки решения. Пример конкретизации данного паттерна информацией о методе когнитивного моделирования приведен на рис. 6.

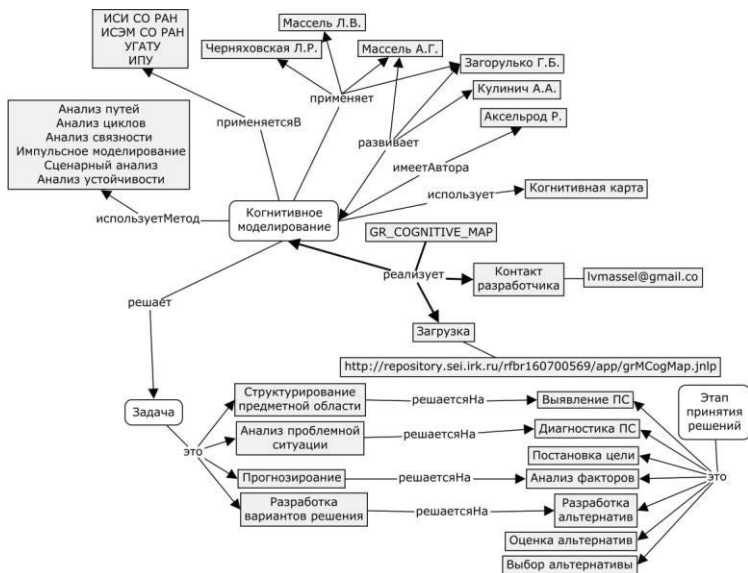


Рис. 6. Конкретизация паттерна содержания «Метод ППР»

Из рис. 6 видно, что метод когнитивного моделирования использует когнитивные карты, а также такие методы анализа путей, анализа циклов, сценарный анализ, предназначен для решения задач структурирования предметной области, анализа проблемной ситуации, разработки вариантов решения и т.п.

## Заключение

Таким образом, наш опыт показал эффективность применения онтологических паттернов содержания при разработке онтологий НПО. Это связано, в частности, с тем, что в онтологии любой НПО, как правило, содержится большое количество типовых фрагментов, которые хорошо описываются паттернами содержания.

Кроме того, использование паттернов содержания позволяет обеспечить единообразное и согласованное представление всех сущностей онтологии НПО, сократить число ошибок онтологического моделирования, повысить «понимаемость» онтологии разработчиками и благодаря этому обеспечить возможность коллективной разработки онтологий.

Благодаря тому, что использование паттернов ОП значительно упрощает и облегчает разработку онтологии НПО, к ней можно привлечь экспертов в моделируемой области, не владеющих навыками онтологического моделирования, что ускоряет разработку онтологии НПО.

## Список литературы

- [Боргест, 2022] Боргест Н.М. Онтология проектирования научного направления: формирование, развитие, примеры // Онтология проектирования. – 2022. – Т. 12, № 2 (44).
- [Грибова и др., 2022] Грибова В.В., Паршкова С.В., Федорищев Л.А. Онтологии для разработки и генерации адаптивных пользовательских интерфейсов редакторов баз знаний // Онтология проектирования. – 2022. – Т. 12, № 2 (44).
- [Загорулько и др., 2016] Загорулько Ю.А., Загорулько, Г.Б., Боровикова О.И. Технология создания тематических интеллектуальных научных интернет-ресурсов, базирующаяся на онтологии // Программная инженерия. – 2016. – Т. 7, № 2.
- [Загорулько и др., 2017] Загорулько Ю.А., Боровикова О.И., Загорулько Г.Б. Применение паттернов онтологического проектирования при разработке онтологий научных предметных областей // Сборник научных трудов XIX Международной конференции DAMDID/RCDL'2017, Москва, Россия. 2017
- [Загорулько и др., 2021] Загорулько Г.Б., Загорулько Ю.А. Применение паттернов онтологического проектирования для разработки онтологии поддержки принятия решений // Труды девятнадцатой Национальной конференции по искусственному интеллекту с международным участием КИИ-2021, Таганрог, Россия. 2021.
- [Загорулько и др., 2021a] Загорулько Ю.А., Сидорова Е.А., Загорулько Г.Б., Ахмадеева И.Р., Серый А.С. Автоматизация разработки онтологий научных предметных областей на основе паттернов онтологического // Онтология проектирования. – 2021. – Т. 11, № 4 (42).
- [Кузнецов и др., 2010] Кузнецов О.П., Суховеров В.С., Шипилина Л.Б. Онтология как систематизация научных знаний: структура, семантика, задачи // Технические и программные средства систем управления, контроля и измерения. Материалы конференции с международным участием. – М.: Институт проблем управления им. Трапезникова РАН, 2010.
- [Кузнецов и др., 2016] Кузнецов О.П., Суховеров В.С. Онтологический подход к оценке тематики научного текста // Онтология проектирования. – 2016. – Т. 6, № 1 (19).
- [Цуканова, 2015] Цуканова Н.И. Онтологическая модель представления и организации знаний: Учебное пособие для вузов. – М.: Горячая линия – Телеком, 2015.
- [Antoniou et al., 2009] Antoniou G., Harmelen F. Web Ontology Language: OWL // Handbook on Ontologies. Staab, S. and R. Studer. (eds). – Berlin: Springer Verlag, 2009.
- [Asim et al., 2018] Asim M.-N., Wasim M., Khan M.U.G. et al. A survey of ontology learning techniques and applications // Database: The Journal of Biological Databases and Curation. – 2018. – Vol. 2018.
- [Blomqvist et al., 2016] Blomqvist E., Hammar K., Presutti V. Engineering Ontologies with Patterns: The eXtreme Design Methodology. Ontology Engineering with Ontology Design Patterns. Studies on the Semantic Web. Hitzler, P., Gangemi, A., Janowicz, K., Krisnadhi, A., Presutti, V. (eds). Amsterdam: – IOS Press, 2016. – Vol. 25.
- [Brusa et al., 2008] Brusa G., Caliusco M. L., Chiotti O. Towards ontological engineering: a process for building a domain ontology from scratch in public administration // Expert Systems. – 2008. – Vol. 25.

- [**DCMI, 2022**] DCMI Metadata Terms. – <http://dublincore.org/documents/dcmi-terms>.
- [**De Nicola et al., 2016**] De Nicola A., Missikoff M. A Lightweight Methodology for Rapid Ontology Engineering // *Com. ACM*. – 2016. – Vol. 59.
- [**Ganino et al., 2018**] Ganino G., Lembo D., Mecella M., Scafoglieri F. Ontology population for open-source intelligence: a GATE-based solution // *Software: Practice and Experience*. – 2018. – Vol. 48, Is. 12.
- [**Gangemi et al., 2009**] Gangemi A., Presutti V. *Ontology Design Patterns // Handbook on Ontologies*, Staab, S., Studer, R. (eds.). – Berlin: Springer Verlag, 2009.
- [**Glinskiy et al., 2020**] Glinskiy B.M., Zagorulko Y.A., Zagorulko G.B., Kulikov I.M., Sapetina A.F., Titov P.A., Zhernyak G.F. Building ontologies for solving compute-intensive problem // International Conference “Marchuk Scientific Readings 2020” (MSR-2020), Novosibirsk, Russia, 2020. *Journal of Physics: Conference Series*. – 2021. – Vol. 1715.
- [**Hammar, 2015**] Hammar K. *Ontology Design Patterns in WebProtégé // Proc. of 14th International Semantic Web Conference (ISWC-2015). Posters & Demonstrations Track. CEUR Workshop Proceedings. Vol. 1486: – URL: [http://ceur-ws.org/Vol-1486/paper\\_50.pdf](http://ceur-ws.org/Vol-1486/paper_50.pdf) (дата обращения: 28.08.2022).*
- [**NeOn Project, 2022**] NeOn Project, 2022. – [http://www.neon-project.org/nw/Welcome\\_to\\_the\\_NeOn\\_Project](http://www.neon-project.org/nw/Welcome_to_the_NeOn_Project).
- [**Noy et al, 2001**] Noy, N., McGuinness D. *Ontology Development 101: A Guide to Creating Your First Ontology // Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.*
- [**ODPA, 2022**] Association for Ontology Design & Patterns. – 2022. – <http://ontologydesignpatterns.org>.
- [**ODPS, 2009**] *Ontology Design Patterns (ODPs) Public Catalog, 2009. – <http://odps.sourceforge.net>.*
- [**Sattar et al., 2020**] Sattar A., Salwana E., Surin M., Ahmad M. N., Ahmad M., Mahmood A.K. Comparative Analysis of Methodologies for Domain Ontology Development: A Systematic Review // *International Journal of Advanced Computer Science and Applications (IJACSA)*. – 2020. – No. 11 (5).
- [**Snytnikov et al., 2020**] Snytnikov A.V., Glinskiy B.M., Zagorulko G.B., Zagorulko Y.A. Ontological approach to formalization of knowledge in computational plasma physics // *Third Virtual Workshop on Numerical Modeling in MHD and Plasma Physics, Novosibirsk, Russia, 2020. Journal of Physics: Conference Series. – 2020. – Vol. 1640.*
- [**Singh et al, 2019**] Singh S, Singh A.M. *Ontology Learning Procedures Based on Web Mining Techniques (March 14, 2019). Proc. of International Conference on Advances in Engineering Science Management & Technology (ICAESMT), Uttaranchal University, Dehradun, India, 2019.*
- [**Suárez-Figueroa et al., 2015**] Suárez-Figueroa M. C., Gomez-Perez A., and Fernández-López M. The NeOn Methodology framework: A scenario-based methodology for ontology development // *Applied Ontology*. – 2015. – Vol. 10.
- [**Zagorulko et al., 2018**] Zagorulko Yu., Borovikova O., Zagorulko G. Pattern-Based Methodology for Building the Ontologies of Scientific Subject Domains // *Proc. of the 17th International Conference SoMeT\_18, Granada, Spain, 2018.*

## КЛАССИФИКАЦИЯ ЭМОЦИЙ ПАЦИЕНТА ПРИ НЕВЕРБАЛЬНОЙ КОММУНИКАЦИИ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ\*

И.С. Косачев (*ilyastalk@bk.ru*)

О.Н. Сметанина (*smoljushka@mail.ru*)

ФГБОУ ВО Уфимский Государственный авиационный  
технический университет, Уфа

Данная статья посвящена решению задачи классификации эмоций пациента при невербальной коммуникации на основе машинного обучения. В статье отражено современное состояние проблемы в области классификации эмоций человека при невербальной коммуникации, приведены постановка задачи и результаты экспериментальных исследований по поиску наиболее качественной модели, позволяющей проводить классификацию эмоций. Полученная модель имеет архитектуру, включающую две рекуррентные подсети с механизмом внимания, выходы которых объединяются и подаются на полносвязный слой классификации. В результате полученная модель на валидационном наборе данных имеет 0.77 ассигасу. Разработка модели велась на языке python с использованием фреймворков tensorflow и keras. Для извлечения изображения лица использовалась модель BlazeFace из фреймворка MediaPipe.

**Ключевые слова:** классификация эмоций человека, невербальная коммуникация, мультимодальные данные, машинное обучение, нейронные сети.

### Введение

Эмоциональное состояние человека является важным показателем, который отражает субъективное отношение человека к различным ситуациям и объектам. Авторы статьи [Холмогорова А.Б. и др., 2019] отмечают, что при госпитализации и наблюдении в стационаре эмоциональное состояние у 22% пациентов определено как высокий уровень тревоги, у 5% – клинически выраженные симптомы тревоги (по шкале А. Бэка). Это в

---

\* Работа выполнена при финансовой поддержке РФ (проект № 22-19-00471).

свою очередь может сопровождаться страхом, грустью и др. эмоциями. При этом пациенту может понадобиться психологическая помощь, а также повлиять на результат лечения. Механизм классификации эмоций может помочь машинам лучше понимать людей и тем самым улучшить качество оказания медицинских услуг.

Человек выражает свои эмоции в процессе вербальной и невербальной коммуникации. Вербальная коммуникация представляет собой способ передачи информации с помощью слов [Станишевская, 2016]. К такому типу коммуникации относится как устная речь, так и письменная. Под невербальными средствами коммуникации понимают способ передачи информации без использования слов. К ним относится мимика, язык телодвижений, голос [Мартынова 2014].

Рэй Бердвистелл обнаружил, что человек передает вербально менее 35% информации, а свыше 65% – невербально [Пиз и др., 2017]. Это означает, что большая часть информации передается с помощью мимики, голоса, языка телодвижений и т. д. Поэтому в статье рассматривается классификация эмоций при невербальной коммуникации.

Вопросам классификации эмоций посвятили свои работы многие российские и зарубежные специалисты, такие как О.П. Тимофеева, С.А. Неимушев, Л.И. Неимущева, И.А. Тихонов [Тимофеева и др., 2020]; Youngsung Kim, ByungIn Yoo, Youngjun Kwak, Changkyu Choi, Junmo Kim [Kim et al., 2017]; Dong Liu, Zhiyong Wang, Lifeng Wang, Longxi Chen [Liu et al., 2021]; Chumachenko K., Iosifidis A., Gabbouj M. [Chumachenko et al., 2022]; Luna-Jimenez C., Kleinlein R., Griol D., Callejas Z., Montero J., Fernandez-Martinez F. [Luna-Jimenez et al., 2021][Luna-Jimenez et al., 2022] и др.

В статье рассматривается современное состояние проблемы, постановка задачи, решение задачи классификации эмоций человека при невербальной коммуникации на основе выбора архитектуры нейросетевой модели и проведении экспериментальных исследований.

Специфической особенностью исследования, результаты которого отражены в данной статье, выступает анализ нескольких модальностей, с помощью которых человек выражает свои эмоции, а именно мимики и голоса.

## **1. Современное состояние проблемы**

О.П. Тимофеева, С.А. Неимушев, Л.И. Неимущева, И.А. Тихонов [Тимофеева и др., 2020] для решения задачи распознавания эмоций используют сверточную нейросеть, которая классифицирует эмоции по изображению лица. В качестве набора данных используется FER-2013, который содержит 7 классов эмоций. Для регуляризации сети они использовали Batch Normalization и Dropout.

Авторами Youngsung Kim, ByungIn Yoo, Youngjun Kwak, Changkyu Choi и Junmo Kim [Kim et al., 2017] решена задача распознавания эмоций по мимике с помощью нейронных сетей. Идея основана на том, чтобы обучить нейросеть классифицировать эмоцию по разнице между эмоциональным выражением лица и обычным. На вход такой модели подается два изображения: лицо с обычным выражением и эмоциональным. Однако, при использовании этой модели на реальных данных, изображение с обычным выражением лица может отсутствовать. Для решения этой проблемы исследователями использована генеративная нейросеть. Она принимает на вход изображение с эмоцией, а на выход генерирует изображение с нейтральным выражением лица.

В статье Dong Liu, Zhiyong Wang, Lifeng Wang и Longxi Chen [Liu et al., 2021] для решения задачи распознавались эмоции по видео и голосу. Для извлечения признаков из видео использована функция Inception-ResNet-v2 приложения Keras. Для извлечения признаков из звука, аудиозапись предварительно проходит предобработку с применением оконного преобразование Фурье, а далее пропускается через комбинацию сверточных и рекуррентных слоёв нейросети: в начале аудио сворачивается, а после полученные признаки подаются на блоки Long Short-Term Memory (LSTM) [Hochreiter et al., 1997].

После извлечения признаки аудио и видео пропускаются через слои LSTM, а после, с помощью метода Хи-квадрат, выделяются самые полезные признаки. В конце признаки аудио и видео объединяются и подаются на вход LIBSVM для классификации.

В качестве наборов данных для обучения авторами [Liu et al., 2021] использованы MOSI и MELD.

Некоторые рассмотренные решения для классификации эмоций используют только мимику. Это может привести к недостоверной классификации, так как человек выражает свою эмоцию не только с помощью выражения лица, но и с помощью голоса, языка телодвижений. Dong Liu, Zhiyong Wang, Lifeng Wang и Longxi Chen [Liu et al., 2021] решают эту проблему. Однако модель для извлечения признаков из изображения лица обучается одновременно с обучением всей модели. Это позволяет лучше находить зависимости в данных, но при этом требует больше вычислительной мощности.

Таким образом, для решения задачи используются различные методы. Самым простым способом является классификация по ключевым точкам лица. Для этого могут использоваться различные классические алгоритмы, такие как решающие деревья и машина опорных векторов.

Для классификации эмоции по изображению лица могут использоваться сверточные нейронные сети, а также генеративные сети [Kim et al., 2017] и STN [Luna-Jimenez et al., 2021]. Для классификации по аудио мо-



гут использоваться сверточные сети [Luna-Jimenez et al., 2021], [Liu et al., 2021], [Chumachenko et al., 2022], рекуррентные сети [Liu et al., 2021], трансформеры [Luna-Jimenez et al., 2022].

Для объединения модальностей могут быть использованы как обычные полносвязные слои [Chumachenko et al., 2022] так и классические алгоритмы, такие как логистическая регрессия [Luna-Jimenez et al., 2022] и машина опорных векторов [Luna-Jimenez et al., 2021].

Более сложным является использование сверточных и рекуррентных нейросетей. Сверточные сети позволяют извлекать признаки из изображения. С их помощью можно получить информацию из изображений лица и спектрограмм. Рекуррентные сети позволяют извлекать информацию из последовательности данных.

Для обучения нейронной сети могут быть использованы ресурсы Google Colaboratory. При необходимости учитывать ограничения, связанные с вычислительными мощностями при обучении модели, предложено разработать собственное программное решение. Такое решение основывается на том, что обучение сети для извлечения признаков из изображения лица и сеть классификации обучаются отдельно.

## 2. Постановка задачи

Разработать модель для классификации объектов, характеризующихся эмоциональной окраской. В качестве входных данных подаётся видео со звуком, в котором присутствует лицо человека. Выходными данными является класс эмоций, включающий его название. Классификация происходит на основе обученной модели нейронной сети, которая принимает на вход элементы невербального общения (голос, последовательность изображений лица).

Математическая постановка задачи может быть сформулирована следующим образом. Дано: множество исходных видеофайлов  $A$ ; множество моделей классификации  $F'$ ; где отображение  $f': a \rightarrow b$ , обеспечивает соответствие между элементами множеств  $A$  и  $B$  ( $b$  - элемент некоторого множества  $B$  (классов)). Обучающая выборка может быть представлена множеством пар  $\{(a, b) | a \in A, b \in B\}$ . Элементы множества  $F''$  позволяют аппроксимировать отображение  $f''$  на полном множестве  $A$ .

Элементы множеств  $f'$  и  $f''$  характеризуются метрикой *accuracy*.

Выявить: метрики ассигасы для  $F''$  с учетом результатов проведенных экспериментов, для выбора наилучшего классификатора.

### 3. Решение задачи классификации эмоций человека при невербальной коммуникации и выбор лучшего классификатора на основе экспериментальных исследований

Методика решения включает три этапа: извлечение аудио и последовательности кадров из видеозаписи, предварительную обработку извлеченных данных и непосредственно классификацию.

Модель классификации должна принимать на вход последовательность признаков лица и мел-спектрограмму. Из каждой модальности извлекаются признаки и объединяются в один вектор. После этот вектор подается на вход слою классификации, который выдает класс эмоции.

Для извлечения признаков из изображения лица используется сверточная нейросеть, обучающаяся отдельно от сети для классификации.

Извлечение признаков из изображения лица человека. Для работы модели классификации нужна нейросеть, которая будет извлекать признаки из изображения лица человека. В качестве такой модели хорошо подходит сверточная нейронная сеть (convolution neural network, CNN). CNN представляет собой модель нейронной сети, которая состоит из сверточных слоёв и преимущественно используется в работе с изображениями, но также имеет применение и в других сферах. Модель предложена в 1998 Яном Лекуном [LeCun et al., 1998]. При разработке модели нейронной сети было решено использовать идею Identity и Convolution блоков ResNet [He et al., 2016] (рис. 1).

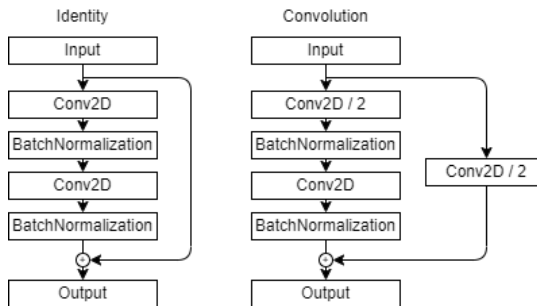
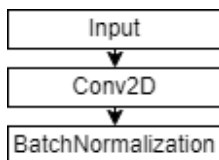


Рис. 1 Архитектура Identity и Convolution блоков

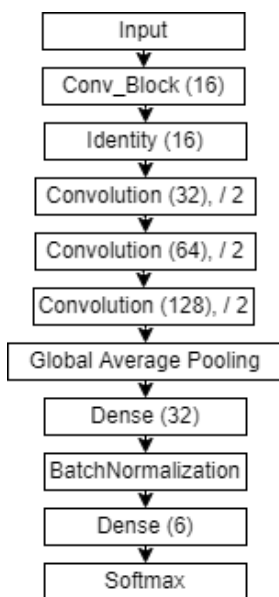
В качестве набора данных для обучения использовался FER-2013. Так как данные в нем сильно несбалансированны, было принято решение убрать из выборки класс *disgusted*. Также, для искусственного увеличения выборки, были добавлены аугментации: случайное отражение изображения по вертикали и случайный поворот изображения от 0 до 25 градусов.

Перед подачей изображений в нейросеть данные проходили предобработку: перевод к оттенкам серого; нормализация; перевод размера изображения к 48x48.

В итоге архитектура сети имеет вид, представленный на рис. 3.



*Рис. 2 Структура Conv\_Block*



*Рис. 3 Архитектура сети*

Сеть обучалась 100 эпох с оптимизатором Adam и с использованием Categorical Crossentropy в качестве функции потерь. Для регуляризации сети использовались слои Dropout.

Оценка модели проводилась на тестовом наборе данных FER-2013. При оценке классификатора использован ряд метрик (рис. 4) и получена следующая матрица ошибок (рис. 5).

	precision	recall	f1-score	support
0.0	0.63	0.59	0.61	958
1.0	0.55	0.39	0.45	1024
2.0	0.85	0.88	0.87	1774
3.0	0.58	0.67	0.62	1233
4.0	0.54	0.56	0.55	1247
5.0	0.77	0.82	0.79	831
accuracy			0.67	7067
macro avg	0.65	0.65	0.65	7067
weighted avg	0.66	0.67	0.66	7067

Рис. 4 Метрики

	angry	fearful	happy	neutral	sad	surprised
<b>angry</b>	566	72	39	105	147	29
<b>fearful</b>	123	395	40	134	228	104
<b>happy</b>	27	23	1565	90	33	36
<b>neutral</b>	60	62	91	824	180	16
<b>sad</b>	104	107	69	253	697	17
<b>surprised</b>	19	55	38	21	17	681

Рис. 5 Матрица ошибок

Из матрицы ошибок видно, что данная модель лучше всего классифицирует класс **happy** и очень часто путает **fearful** с классом **sad**.

Для извлечения признаков из изображения лица все слои после Global Average Pooling 2D убираются.

*Данные для обучения.* На вход модель классификации принимает видео и аудио. Авторами использован набор данных RAVDESS [Livingstone et al., 2018]. Причины выбора обусловлены следующим: набор данных содержит видео и аудио; все видео четко разделены по фразам и эмоциям; в каждом видео присутствует только один актер, а значит не нужно решать проблему с определением того, кто говорит в данный момент.

Набор данных содержит 8 меток классов: **neutral**, **calm**, **happy**, **sad**, **angry**, **fearful**, **disgust**, **surprised**. Для обучения классы **neutral** и **calm** были объединены.

Перед подачей в нейросеть аудио и видео проходят предобработку: разделение видео на последовательность изображений; извлечение из изображений лица человека, с помощью BlazeFaceMediaPipe [Bazarevsky et al., 2019]; перевод изображений в оттенки серого, нормализация и изменение размера изображения до 48x48; извлечение признаков из изображений лица; изменение размера полученного тензора до 150x128 (обрезая или наоборот заполняя нулями).

Предобработка звука включает следующие шаги: приведение частоты дискретизации к 16 КГц; приведение к одноканальному формату; извлечение мел-спектрограммы; изменение размера мел-спектрограммы до 100x128 (обрезая или заполняя нулями).

Выбор архитектуры модели для классификации эмоций по голосу и мимике на основе экспериментальных исследований. Во всех экспериментах в качестве функции потерь использовалась Categorical Crossentropy, в качестве оптимизатора Adam с шагом обучения 0.001 и уменьшением его на 0.1 каждые 20 эпох. Всего каждая модель обучалась 60 эпох. Перед обучением выборка была разделена обучающую и валидационную в соотношении 70/30, причем актеры из валидационной выборки не встречаются в обучающей. Во время обучения модели сохранялась конфигурация наи-

лучшей модели, которая определялась по наименьшим потерям на валидационной выборке. Рассмотрены три вида архитектур нейронной сети для классификации эмоций по голосу и мимике (рис. 6): VGG\_RNN (а), RNN\_RNN (б), A\_V\_RNN (в).

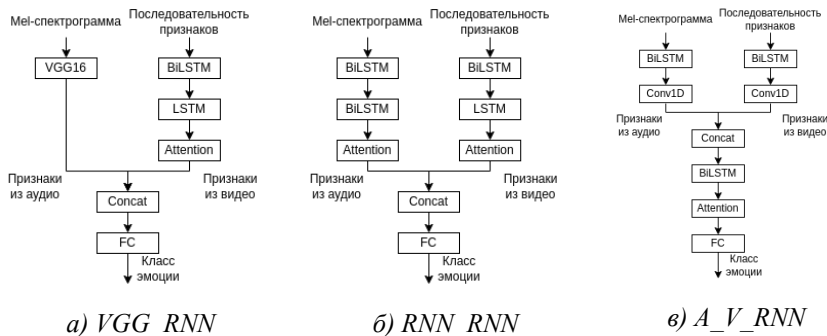


Рис. 6. Архитектуры нейронной сети для классификации эмоций по голосу и мимике

Первая архитектура имеет название VGG\_RNN (рис. 6,а). В ней для извлечения информации из последовательности признаков лица используются слои BiLSTM, LSTM и concatAttention. Для извлечения признаков из мел-спектрограммы используется сверточная нейросеть VGG16. Эксперименты проводились с дообучением VGG16. В первом эксперименте обучались только 4 слоя VGG16, во втором – все слои. Результаты обучения модели представлены на рис. 7–10.

	precision	recall	f1-score	support
0	0.84	0.72	0.78	80
1	0.60	0.85	0.70	120
2	0.62	0.62	0.62	40
3	0.64	0.71	0.67	80
4	0.81	0.62	0.70	80
5	0.76	0.68	0.72	80
6	0.22	0.10	0.14	40
accuracy			0.67	520
macro avg	0.64	0.62	0.62	520
weighted avg	0.67	0.67	0.66	520

Рис. 7. Метрики для VGG\_RNN (4 слоя VGG16)

	angry	calm	disgust	fearful	happy	sad	surprised
angry	58	11	8	3	0	0	0
calm	1	102	1	0	8	0	8
disgust	1	5	25	0	0	9	0
fearful	5	10	2	57	1	5	0
happy	0	16	0	8	50	0	6
sad	0	12	4	8	2	54	0
surprised	4	15	0	13	1	3	4

Рис. 8. Матрица ошибок для VGG\_RNN (4 слоя VGG16)

	precision	recall	f1-score	support
0	0.88	0.72	0.79	80
1	0.69	0.81	0.75	120
2	0.52	0.68	0.59	40
3	0.64	0.69	0.66	80
4	0.87	0.65	0.74	80
5	0.65	0.80	0.72	80
6	0.22	0.10	0.14	40
accuracy			0.69	520
macro avg	0.64	0.64	0.63	520
weighted avg	0.68	0.69	0.68	520

Рис. 9. Метрики для VGG\_RNN (все слоу VGG16)

	angry	calm	disgust	fearful	happy	sad	surprised
angry	58	8	9	5	0	0	0
calm	0	97	3	2	6	4	8
disgust	0	4	27	0	0	9	0
fearful	6	2	6	55	0	11	0
happy	0	12	0	9	52	1	6
sad	0	8	6	2	0	64	0
surprised	2	9	1	13	2	9	4

Рис. 10. Матрица ошибок для VGG\_RNN (все слоу VGG16)

Дообучение всех слоёв VGG16 имеет более высокие значения метрик, чем при дообучении только последних четырех. По матрицам ошибок видно, что модель очень плохо справляется с классификацией класса surprised.

Вторая рассмотренная архитектура имеет название RNN\_RNN (рис. 6,б). Для получения информации из признаков лица используется та же архитектура, но для извлечения признаков из мел-спектрограммы в ней вместо VGG16 используются две BiLSTM + concatAttention.

Результаты обучения данной архитектуры представлены рис. 11, 12.

	precision	recall	f1-score	support
0	0.77	0.88	0.82	80
1	0.73	0.90	0.81	120
2	0.74	0.62	0.68	40
3	0.79	0.72	0.76	80
4	0.90	0.75	0.82	80
5	0.82	0.66	0.73	80
6	0.65	0.70	0.67	40
accuracy			0.77	520
macro avg	0.77	0.75	0.75	520
weighted avg	0.78	0.77	0.77	520

Рис. 11. Метрики для RNN\_RNN

	angry	calm	disgust	fearful	happy	sad	surprised
angry	70	5	2	3	0	0	0
calm	0	108	0	0	5	0	7
disgust	4	5	25	0	0	6	0
fearful	11	1	2	58	0	6	2
happy	0	9	0	5	60	0	6
sad	3	16	5	1	2	53	0
surprised	3	3	0	6	0	0	28

Рис. 12. Матрица ошибок для RNN\_RNN

Обучение данной модели дало значительно большие значения метрик, чем при обучении VGG\_RNN. Данная модель значительно лучше классифицирует как класс surprised, так и все остальные классы.

Третья архитектура имеет название A\_V\_RNN (рис. 6,в) и является модификацией второй. Её идея заключается в том, чтобы объединить признаки из спектрограммы и из изображения, а после отправить их на вход LSTM. Для того, чтобы объединить признаки, они сворачиваются одномерной сверткой к одинаковой размерности.

Результаты эксперимента представлены на рис. 13, 14.

	precision	recall	f1-score	support
0	0.81	0.72	0.76	80
1	0.66	0.95	0.78	120
2	0.73	0.68	0.70	40
3	0.89	0.70	0.78	80
4	0.89	0.72	0.80	80
5	0.75	0.53	0.62	80
6	0.70	0.93	0.80	40
accuracy			0.75	520
macro avg	0.77	0.75	0.75	520
weighted avg	0.77	0.75	0.75	520

Рис. 13 Метрики для A\_V\_RNN

	angry	calm	disgust	fearful	happy	sad	surprised
angry	58	11	3	3	1	1	3
calm	0	114	0	0	5	0	1
disgust	4	2	27	0	0	7	0
fearful	5	4	5	56	1	6	3
happy	0	12	0	1	58	0	9
sad	4	30	2	2	0	42	0
surprised	1	1	0	1	0	0	37

Рис. 14 Матрица ошибок для A\_V\_RNN

Третья архитектура имеет ассурасу меньший, чем у второй, и больший, чем у первой. По матрице ошибок видно, что модель плохо классифицирует эмоцию sad.

Обобщающие результаты экспериментов приведены в табл. 1.

Таблица 1

	VGG_RNN (4 слоя VGG16)	VGG_RNN (все слои VGG16)	RNN_RNN	A_V_RNN
Accuracy	0.67	0.69	0.77	0.75
MacroAvg Precision	0.64	0.64	0.77	0.77
WeightedAvg Precision	0.67	0.68	0.78	0.77
MacroAvg Recall	0.62	0.64	0.75	0.75
WeightedAvg Recall	0.67	0.69	0.77	0.75
MacroAvg F1	0.62	0.63	0.75	0.75
WeightedAvg F1	0.66	0.68	0.77	0.75

Среди трех рассмотренных архитектур нейронных сетей самой лучшей оказалась RNN\_RNN, которая для получения информации из аудио и видео использует рекуррентные подсети.

#### 4. Сравнительный анализ решений

Для сравнительного анализа были выбраны мультимодальные модели, которые обучались на датасете Ravdess. Авторы статьи [Luna-Jimenez et al., 2021] классифицировали эмоции с помощью моделей CNN14 из фреймворка PANNs для аудио и STN + Bilstm + attention для видео. В за-

ключение эксперимента проведена классификация по объединенным выходам моделей по аудио и видео с помощью SVM. В результате ими получено значение метрики accuracy – 80.08%.

Модель других авторов (Intermediate-Attention-Fusion) [Chumachenko et al., 2022] имеет схожесть с моделью, предложенной авторами данной статьи. Она заключается в том, что есть отдельная модель для извлечения признаков из изображения (EfficientFace) и отдельная модель для классификации. Также схожестью является то, что классификатор имеет отдельные подсети для извлечения признаков из аудио и видео, выходы которых объединяются и подаются на полносвязный слой. Однако отличием является то, что для извлечения признаков авторы [Chumachenko et al., 2022] используют одномерные свертки, а не рекуррентные сети. Помимо одномерных сверток в модели используются блоки Attention. В результате авторы [Chumachenko et al., 2022] с помощью своей модели получили значение метрики accuracy – 81.58%.

Третья модель разработана авторами [Luna-Jimenez et al., 2022]. В ней для классификации по аудио использованы wav2vec2 + MLP, а для видео вначале извлечены Action Units с помощью OpenFace, а после произведена классификация с помощью рекуррентной Bilstm. В конце объединены выходы двух моделей и произведена классификация с помощью логистической регрессии. В результате получено значение метрики accuracy – 86.7%.

Сравнительные результаты применения мультимодальных моделей приведены в табл. 2.

Таблица 2

Модель	Accuracy
Logistic regression на признаках из Wav2vec2+MLP и ActionUnits+Bilstm+Attention	86.7%
Intermediate-Attention-Fusion	81.58%
SVM на признаках из CNN14 и STN+Bilstm+Attention	80.08%
Разработанная модель	77 %

Авторы статьи в перспективе займутся усовершенствованием мультимодальных моделей для повышения качества распознавания.



## Заключение

Особенностью проведенного исследования является классификация эмоций при невербальной коммуникации человека по двум модальностям, а именно мимике и голосу.

В качестве инструментов разработки используется язык python с использованием библиотек tensorflow + keras, для обучения модели, и фреймворк MediaPipe, для извлечения лиц из изображения. Для разработки нейросетевой модели использовались полносвязные, сверточные и рекуррентные слои.

В ходе исследования проведены эксперименты для выявления наилучшей модели для классификации эмоций. Результаты исследований показали, что модель имеет архитектуру, состоящую из двух рекуррентных подсетей, выходы которой объединяются и подаются на полносвязный слой классификации.

В результате обучения модели на валидационной выборке были получены следующие значения метрик: 0.77 Accuracy, 0.77 MacroAvg Precision, 0.78 WeightedAvg Precision, 0.75 MacroAvg Recall, 0.77 WeightedAvg Recall, 0.75 MacroAvg F1, 0.77 WeightedAvg F1. Данные значения показывают, что модель имеет хорошее качество классификации.

В перспективе, для улучшения качества классификации, можно добавить ещё несколько модальностей, например, позу тела человека; собрать свой набор данных, который содержал бы записи на русском языке; попробовать использовать вместо мел-спектрограмм мел-кепстральные коэффициенты.

Предложенное решение может быть использовано во многих сферах: в медицине может помочь лучше определять психологические проблемы, а также улучшить автоматизацию ухода за пациентами; в образовательной сфере для адаптации заданий для обучения и мониторинга эмоционального состояния учащихся; в различных организациях также может использоваться для мониторинга эмоций работников, чтобы не допускать выгорания сотрудников; в индустрии разработки видеоигр может быть использовано для изменения окружения в зависимости от эмоции игрока.

## Список литературы

[Холмогорова А.Б. и др., 2019] Холмогорова А.Б., Суботич М.И., Рахманина А.А., Борисоник Е.В., Рой А.П., Лысенко Ю.С., Петриков С.С. Испытываемый уровень стресса и симптомов у пациентов многопрофильного медицинского центра // – Журнал им. Н.В. Склифосовского «Неотложная медицинская помощь». – 2019. – 8 (4). – 384-390. – <https://doi.org/10.23934/2223-9022-2019-8-4-384-390>.

- [Пиз и др., 2017] Пиз Алан, Пиз Барбара. Новый язык телодвижений. – М.: Эксмо, 2017.
- [Мартынова, 2014] Мартынова Е.М. Невербальная коммуникация: теории и мнения // Вестник ЮУрГГПУ. – 2014. – № 8.
- [Станишевская, 2016] Станишевская Ж. Вербальная коммуникация как объект исследования в процессе психотерапии // Studia Humanitatis. – 2016. – № 2.
- [Тимофеева и др., 2020] Тимофеева О.П., Неимущев С.А., Неимущева Л.И., Тихонов И.А. Распознавание эмоций по изображению лица на основе глубоких нейронных сетей // Труды НГТУ им. Р.Е. Алексеева. – Нижний Новгород, 2020. – № 1 (128).
- [Bazarevsky et al., 2019] Bazarevsky V., Kartynnik Y., Vakunov A., Raveendran K., Grundmann M. BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs // arxiv. 2019.
- [Chumachenko et al., 2022] Chumachenko K., Iosifidis A., Gabbouj M. Self-attention fusion for audiovisual emotion recognition with incomplete data // arxiv. 2022.
- [Liu et al., 2021] Liu D., Wang Z., Wang L., Chen L. Multi-Modal Fusion Emotion Recognition Method of Speech Expression Based on Deep Learning // Frontiers in Neurorobotics. – 2021. – Vol. 15.
- [Luna-Jimenez et al., 2021] Luna-Jimenez C., Kleinlein R., Griol D., Callejas Z., Montero J., Fernandez-Martinez F. A proposal for Multimodal Emotion Recognition using aural transformers and action units on RAVDESS dataset // Applied Sciences. – 2022. – Т. 12, No. 1.
- [Luna-Jimenez et al., 2021] Luna-Jimenez C., Kleinlein R., Griol D., Callejas Z., Montero J., Fernandez-Martinez F. Multimodal emotion recognition on RAVDESS dataset using transfer learning // Sensors. – 2021. – Vol. 21, No. 22.
- [Hochreiter et al., 1997] Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural computation. – 1997. – Vol. 9, No. 8.
- [He et al., 2016] He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // IEEE Conference on Computer Vision and Pattern Recognition. – 2016.
- [LeCun et al., 1998] LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition. // Proceedings of the IEEE. – 1998. – Vol. 86, No. 11.
- [Livingstone et al., 2018] Livingstone S., Russo F. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English // PLoS ONE. – 2018. – Vol. 13, No. 5.
- [Kim et al., 2017] Kim Y., Yoo B., Kwak Y., Choi C., Kim J. Deep generative-contrastive networks for facial expression recognition // arXiv. 2017.

## АВТОМАТИЗИРОВАННАЯ ТЕКСТОВАЯ ПСИХОДИАГНОСТИКА И ПРОБЛЕМЫ МОНИТОРИНГА СОЦИАЛЬНЫХ СЕТЕЙ

Ю.М. Кузнецова (*kuzjum@ya.ru*)

Н.В. Чудова (*nchudova@gmail.com*)

А.А. Чуганская (*anfi.chuganskaya@yandex.ru*)

ФИЦ ИУ РАН, Москва

В работе представлен обзор результатов текстовой автоматизированной психодиагностики, полученных с помощью созданного в ФИЦ ИУ РАН инструмента интеллектуального анализа текста. Предложен теоретический анализ проблем организации психодиагностического мониторинга социальных сетей, показавший, что требуется как решение ряда теоретических вопросов социальной психологии больших групп, так и совершенствование методов интеллектуального анализа текста, которые позволили бы ставить задачу организации изучения эмоциональных реакций пользователей на социально значимые события.

**Ключевые слова:** психолингвистика, интеллектуальный анализ текста, сетевое общение, мониторинг социальных сетей.

### Введение

Задачи исследования и управления в социо-гуманитарной сфере на современном этапе развития технологий определяют необходимость разработки инструментов мониторинга эмоциональных реакций и психологических особенностей пользователей социальных сетей. Интерес к измерению психоэмоционального состояния пользователей социальных сетей на основе данных интеллектуального анализа их текстовых сообщений приводит к необходимости проводить исследования, посвящённые проблеме взаимосвязи речи и психологических особенностей авторов текстов.

## 1. Психолингвистические исследования личностных особенностей

В мировой науке на современном этапе сформировалась область междисциплинарных исследований, направленных на обнаружение связей между психологическими особенностями человека и характеристиками порождаемой им текстовой продукции. Особый интерес для исследователей представляют собой тексты сетевого общения как источник психодиагностической информации индивидуального, группового и популяционного уровней. Совершенствованию в автоматическом выявлении маркеров личностных качеств и эмоционального состояния, а также уточнению текстовых признаков психического неблагополучия пользователей соцсетей посвящены систематически проводимые международные соревнования программных средств. В инструментах, создаваемых программистами и компьютерными лингвистами, решение задач психодиагностики основывается преимущественно на лексических и морфологических характеристиках текста.

В психодиагностике базой для проведения психолингвистических исследований прикладного характера является представление о том, что уникальные речевые паттерны отражают устойчивые индивидуальные характеристики человека, такие, как черты личности или особенности эмоционального реагирования. С самого начала в исследованиях, посвященных связям между речью / текстом и характеристиками говорящего / автора, первостепенное внимание уделялось поиску лингвистических признаков психопатологических состояний. В работах основоположников данного направления Е. Bleuler, К. Kleist, Е. Краепelin, П.И. Карпова, В.Н. Образцова, В.П. Осипова и др. были описаны диагностически значимые формы искажений, возникающих в речевой деятельности под влиянием психических заболеваний. Было выявлено, что наиболее общим дефектом, сопровождающим развитие психической патологии, является снижение смысловой и формальной связности высказываний (феномены атаксии и «разорванной речи») [Микиртумов, и др., 2012]. В дальнейшем применение количественных и статистических методов анализа позволило ввести в терапевтическую и исследовательскую практику ряд вычисляемых индексов, отражающих специфические особенности употребления некоторых частей речи, а также семантики и структуры высказывания (коэффициент Трейгера, коэффициент определенности действия, показатель Готтшалка-Глезера, индекс прегнантности Эртеля, индекс Шлисмана, фог-индекс и др.) [Белянин, 2003].

Помимо исследований речевых проявлений психопатологии психолингвистические средства традиционно применяются для выявления признаков личностных особенностей участников коммуникации. Часто в ка-

честве психологических характеристик в таких исследованиях выступают показатели факторных психодиагностических методик. К настоящему времени накоплен большой объем информации о представленности в текстах на разных языках информации, свидетельствующей о выраженности черт личности, измеряемых с помощью методики «Большая пятерка» [El-Demerdash et al., 2020; Moreno et al., 2021; Tinwala et al., 2021] и ряда других опросников. Так, например, в исследовании с использованием русской адаптации «Большой пятерки» и Фрайбургского многофакторного личностного опросника FPI описаны такие связи, как отрицательные корреляции для шкал «Экстраверсия» и «Доброжелательность» с долей слов длиннее 6 букв, шкалы «Добросовестность» с долей слов, обозначающих эмоции; выявлено, что уровень невротичности и депрессивности положительно коррелирует с долей местоимений и долей личных местоимений; а со шкалой эмоциональной лабильности положительно коррелирует доля союзов и отрицательно – доля слов тематической группы «Пространство» [Литвинова и др., 2015].

С помощью психолингвистических методов в коммуникации могут выделяться признаки, образующие психолингвистический профиль (портрет), который включает в себя такие восстанавливаемые по тексту компоненты, как демографический профиль коммуниканта (возраст, пол, социальный статус, раса и т.п.), а его также характерологические, эмоциональные и мотивационные особенности [Образцов и др., 2002; Gjurković et al., 2020].

Возможность оперировать большими объемами текстовых данных стала фактором модернизации не только предмета, но и методов прикладной психолингвистики [Загоровская и др., 2016].

## **2. Реляционно-ситуационный анализ текста как основа построения семейства методов текстовой автоматизированной психодиагностики**

Метод реляционно-ситуационного анализа (РСА) основан на теории коммуникативной грамматики русского языка, разработанной в ИРЯ РАН, и на теории неоднородных семантических сетей, созданной в ФИЦ ИУ РАН [Осипов и др., 2008]. Метод широко применяется в задачах поиска релевантной информации, сравнения документов и классификации коллекций текстов [Осипов и др., 2016]. Особенно актуально использование подобных инструментов в связи с необходимостью анализа данных в рамках изучения больших социальных групп и, соответственно, связанных с этим данных (Big Data) в сетевом пространстве. В последние годы метод РСА стал применяться для выявления текстовых признаков психологических особенностей авторов, в том числе многочисленных пользователей социальных сетей. Связано это с тем, что метод РСА позволяет выявлять

предикатно-аргументативную структуру высказываний на естественном языке, что согласуется с психологическими представлениями (в частности, отечественной школы Выготского-Лурии) об особой роли глаголов в разворачивании мысли говорящего/пишущего.

Совершенствование метода [Smirnov et al., 2021] происходило как за счет разработок в области автоматического установления семантических ролей [Shelmanov et al. 2014] и создания новых алгоритмов, работающих на PCA-анализаторе [Devyatkin et al., 2020; Devyatkin et al., 2021], так и за счёт психолингвистических и лингвистических работ [Григорьев и др., 2022; Никитина и др., 2022].

Разработанный инструментарий использовался в нескольких психологических исследованиях, позволивших установить ряд психолингвистических закономерностей.

Выделены текстовые признаки, коррелирующие с результатами психодиагностического исследования и/или экспертной оценки:

- -агрессивности как личностной черты: повышение частотности в текстах лексики социального разобщения, страдания, аффектации и напряжения, прилагательных, высокая частота встречаемости семантической роли «деструктив», глаголов первого лица, большое количество знаков препинания, причастий, деепричастий;
- отсутствия эмпатии как компонента нарциссизма: высокая частотность лексем, относящихся к категории безысключительной и усиленной лексики, лексики отрицательной рациональной оценки, лексики социальной разобщенности; коэффициент опредмеченности действия; коэффициент Трейгера; высокая частотность всех местоимений и личных местоимений первого и третьего лица; частотность глаголов 1 лица ед. числа в прошедшем времени; средняя длина слов;
- актуального на момент написания текста состояния фрустрации: повышение количества местоимений первого лица, знаков препинания, словоформ с отрицательными приставками (*ненадежный, бестолково, нигде* и т.п.), слов с семантикой аффектации (*чудовищный, прекрасно, счастье, офигеть* и т.п.), инвектив, склонность строить высказывания, содержание которых связано с выявлением причин, указанием на объект реального или потенциального разрушения или ликвидации (синтаксемы: каузатив, ликвидатив, деструктив);
- клинической депрессии (пациенты НЦПЗ) и депрессивности как личностной особенности, обнаруживаемой у здоровых людей шкалой Бека. В ходе исследования были проанализированы около 400 текстов (объемом не менее 1 страницы), среди них 239 текстов пользователей социальных сетей с повышенной депрессивностью по шкале Бека и 93 текста пациентов с подтвержденным диагнозом

депрессивного расстройства, проходящих лечение в рамках стационара. С помощью применения к данным инструмента алгоритма установления каузальных связей AQ-JSM было выделено 16 лингвистических маркеров для клинической депрессии и 27 маркеров для депрессивного состояния пользователей соцсетей.

При исследовании статистики текстовых параметров как средства изучения психических процессов и состояний, содержания сознания и картины мира, были выявлены следующие факты:

- предикатно-ролевые структуры отражают на вербальном уровне устройство компонентов или «участков» картины мира субъекта, обычно не становящиеся предметом наблюдения. Так, в текстах эссе «Я. Другие. Мир», выявлены тенденции, характеризующие влияние психического отклонения: частота употребления объекта «Я» в семантической роли «авторизатор» (субъект оценки, восприятия, речи-мысли) в текстах испытуемых из группы психически больных (шизофрения и депрессия) встречается в два раза реже, чем в текстах здоровых, а в роли «адресат» (лицо, к которому обращено информативное, донативное или эмотивное действие) – в два раза чаще;
- существует два типа текстовой объективации психологического барьера, характеризующего состояние фрустрации: автор, относящийся к типу «деятеля», описывает преимущественно свои собственные действия, а автор, относящийся к типу «наблюдателя», склонен к описанию обстоятельств;
- содержание циркулирующего в современном дискурсивном пространстве Рунета «образа старости» является культуроспецифичным: для участников сетевого общения пожилые люди отличаются от представителей других возрастов не своей погруженностью в негативную аффективность, а тем, что кажутся менее способными испытывать и переживать эмоции; выявлена также специфическая деформация гендерных стереотипов эмоциональности, в силу которых она атрибутируется пожилым мужчинам в большей степени, чем пожилым женщинам.
- отклик на либеральную и государственническую пропаганду различается тем объёмом внимания, который отводится эмоциям – своим, собеседников, значимых для общества фигур, социальных групп: в первом случае люди скорее выражают свои эмоции, чем обсуждают их, во втором – больше говорят о своих и чужих переживаниях, обсуждая причины, их вызвавшие; в дискуссиях, стимулированных оппозиционно-либеральной пропагандой, эмоции и их причины реже, чем в неполитических дискуссиях, становятся предметом внимания.

Дальнейшее совершенствование инструмента требует, во-первых, расширения перечня диагностируемых по текстам особенностей их авторов, во-вторых, развития словаря предикатных слов, в первую очередь, расширение круга эмотивов, в-третьих, реализации модульной концепции, которая позволит обеспечить адаптивность инструмента при решении различающихся по предмету исследовательских задач.

В целом, результаты наших эмпирических исследований позволяют выделить в качестве основных перспективных направлений применения разрабатываемого метода групповую диагностику психологических особенностей и психического неблагополучия, а также мониторинговые мероприятия, направленные на определение уровня представленности признаков неблагополучия в сетевом контенте и локализации «горячих точек» коммуникативного пространства. В силу статистической природы выявляемых с помощью нашего инструмента признаков очевидно, что адекватность выводов обеспечивается объемом обрабатываемой текстовой информации, поэтому наилучшие результаты будут демонстрировать мониторинговые исследования.

### **3. Проблемы организации психодиагностического мониторинга социальных сетей**

Среди задач мониторинга соцсетей выделяют три крупные области – задачи поддержки маркетинга, госуправления и научного исследования. В первом случае более востребован оказывается подход Data Science, а в третьем – подход классического ИИ. Связано это с тем, что в первом случае целью проводимой работы с данными является разделение встретившихся случаев на заданные классы, а в третьем – выявление закономерностей, реализующихся в существовании задаваемых классов. Второй же случай, случай использования данных наблюдения в работе управленцев различных уровней, сам по себе неоднороден и, видимо, допускает использование обоих подходов к обработке данных.

Различие между задачами маркетинга и задачами научного исследования можно описать как различия между проблемами, решаемыми практическим и теоретическим мышлением. Согласно определению Б.М.Теплова [Теплов, 1961] практическое мышление – это мышление, регулирующее действие. Теплов выделил ряд критериев, позволяющих отнести конкретную проблему к практическому или теоретическому типу: 1) требуется решение конкретного вопроса – требуется нахождение общей закономерности; 2) проблема должна быть решена за отведённое на это время или просто как можно скорее – проблема должна решаться столько, сколько нужно для нахождения принципа решения; 3) проблему «ставит жизнь» – проблема возникает в результате выполнения процедуры «постановки



проблемы» (компульсивное и произвольное принятие проблемы). Среди задач управления могут встречаться как практические проблемы, так и теоретические. В тех случаях, когда требуется совершить управленческое воздействие, выбираемое из числа известных на основе быстрой категоризации происходящего, управленец имеет дело с практической проблемой и не слишком нуждается в интерпретируемости, «прозрачности» процедуры отнесения конкретных событий к тому или иному классу. Фактически, здесь речь идёт о переименовании – категоризовать происходящее (например, реакцию населения на сообщение о техногенной катастрофе или появление в подростковом сообществе деструктивного увлечения) означает выбрать предписанное инструкцией действие. В том же случае, когда требуется сформировать представление о происходящем как уникальном событии – происходящем с этими людьми в этих обстоятельствах, здесь-и-сейчас – управленец фактически решает творческую задачу. При этом отнесение данного «случая» к тому или иному типу приобретает иной смысл – смысл характеристики объекта по ряду существенных признаков, т.е. смысл познавательного действия.

В целом, всё многообразие ситуаций использования данных по своим интеллектуальным целям может быть сведено к двум ситуациям, описанным М.Вертгеймером – решение «лабиринтной» задачи и решение задачи, допускающей понимание [Вертгеймер М. Продуктивное мышление. 1987.]. Соответственно, и подходов к обработке данных мониторинга соцсетей, в частности, к процедурам анализа текста, существует два – подход «чёрного ящика» и модельный подход. В обработке данных эти две «идеологии» существуют в данный момент как нейросетевой подход (см., например, [Себрант, 2022]) и подход классического ИИ [Поспелов, 1996]. В 21 веке с появлением новых вычислительных возможностей история нейросетевой метафоры, для психологии возникшей в конце 19в, повторилась в работах по ИИ: возникла проекция модели нервной ткани на целый ряд практик и прикладных областей знания, в первую очередь, на прикладные исследования потребительского поведения, диагностические процедуры в медицине, в искусствоведческой или почерковедческой экспертизе и проч. Моделью устройства предмета исследования, как и в психологии ассоцианизма, стала модель целого как скопления гомогенных элементов, связанных между собой по законам сходства и смежности. Но элементами теперь оказываются люди – покупатели, больные, художники, избиратели, люди, мечтающие вступить в брак и т.д. Однако специалистам-предметникам трудно принять замену предмета исследования объектом реальности, в которой он воплощён. Поэтому, параллельно с расширением круга заказчиков, у «современного ИИ» растёт и число противников (см., например, интервью, данное в июле 2022г директором ИПУ РАН Д.А.Новиковым [Новиков, 2022]).

В самом общем виде цели мониторинга соцсетей в интересах управления (государственного или корпоративного) могут быть описаны как выявление «влиятельных» блогеров, выявление групп риска (по психическим или поведенческим девиациям) и выявление реакций пользователей на социально значимые события. Главной проблемой, возникающей при формулировании конкретных целей мониторинга, оказывается проблема выборки. У этой проблемы есть два аспекта – кого изучать и в каком объёме. Вопрос об объекте наблюдения решается, исходя из криминологических, социологических, социально-психологических и психологических приоритетов специалистов, готовящих мониторинг; вопрос о необходимом объёме данных наблюдения связан с первым, т.е. с предметной ориентацией специалистов, проводящих исследование.

Сетевой контент может быть рассмотрен как материал для принятия решений органами правопорядка. Именно созданию инструментов выявления случаев нарушения законов РФ посвящено множество работ по обнаружению т.н. деструктивного контента [Остапенко и др., 2022]. Такая цель выявления фактов противоправной деятельности (разжигание розни, пропаганда нацизма, призывы к терроризму и проч.) не требует анализа социологических, политологических, психологических закономерностей и потому у специалистов нет нужды создавать системы интеллектуального анализа текста, беспокоиться о репрезентативности выборки и работать над обеспечением валидности измеряемых показателей. Достаточно создать поиск, по ключевым словам, и базу данных, в которой фиксируются данные мониторинга в форме, удобной для последующего ручного анализа экспертами. В этой области, как и в системах поддержки маркетинга, применяются методы оценки позитивно-го/негативного отношения к объекту сетевого обсуждения.

Для социолого-ориентированных специалистов при анализе сетевого контента в центре внимания оказываются реакции пользователей как представителей общественного мнения той или иной социальной группы – возрастной, профессиональной, локальной/региональной, экономической и проч. Объём выборки в таком случае задаётся доступностью аккаунтов пользователей, содержащих необходимую биографическую информацию. Это увеличивает объём выкачиваемых постов и комментариев, поскольку далеко не все аккаунты содержат полный объём необходимой информации, но хотя бы имеет понятные для социологов численные критерии репрезентативности выборки.

Для социально-психологических и психологических исследований в сети вопрос о данных мониторинга не может решаться с опорой на социологические представления о генеральной совокупности и репрезентативности выборки. В науке не существует теоретических представлений, позволяющих вычислить репрезентативность для таких объектов как группа (сетевое сообщество) или индивид (пользователь) с определёнными пси-

хологическими (в том числе, патопсихологическими) особенностями. Зато именно такие исследования оперируют не мнениями пользователей, а диагностируемыми на основе данных о сетевом поведении и текстовой продукции особенностями эмоционального и рационального (проблемно-ориентированного) реагирования. Говорить, например, о росте стрессовых реакций или о снижении готовности к совместным усилиям, возможно не на основе высказанных пользователями мнений по тем или иным поводам, а лишь на основе психодиагностической квалификации сказанного/сделанного пользователями.

Изучение групповой динамики в сети может позволить установить пропорцию сплочённых и аморфных сообществ, обсуждающих ту или иную тематику (региональную или общедодедеральную). Социально-психологическая ориентация специалиста, проводящего мониторинг, позволяет анализировать «свойства коллективного субъекта, которые могут иметь разный уровень развития, позволяют раскрыть механизмы формирования и функционирования сетевого сообщества как группы людей, взаимодействующих в дискурсивном пространстве социальных медиа, объединённых особыми связями и отношениями и способных к проявлению совместных форм активности и саморефлексии» [Воронин и др., 2019].

Исследование эмоциональных реакций пользователей и их психологических особенностей представляет, по-видимому, наибольший интерес для практики - во всяком случае, если судить об этом по публикациям в открытой печати, посвящённых развитию методического инструментария. И здесь проблема интерпретации данных анализа сетевого поведения и анализа сетевых дискуссий оказывается центральной. Как уже было сказано, никаких теоретических оснований для определения величины выборки для психологических, а не социологических характеристик человека не существует. Нет и теоретически обоснованного отношения к пользователю соцсетей как к «представителю» социальной группы или всего социума – ответ на вопрос представляет ли пользователь кого-то кроме себя самого зависит от убеждений (часто неререфлексируемых, имплицитных) самого исследователя. Более того, нет и единой модели характеристики социального на основе индивидуально-психологического.

## **Заключение**

Результаты, полученные с помощью созданного в ФИЦ ИУ РАН, позволяют нам ставить задачи организации мониторинга эмоциональных реакций пользователей социальных сетей. Проведённый теоретический анализ проблем организации психодиагностического мониторинга показал, что требуется как решение ряда теоретических вопросов социальной психологии больших групп, так и совершенствование методов интеллектуального анализа текста.

## Список литературы

- [**Белянин, 2003**] Белянин В.П. Психолингвистика. – М.: Флинта Моск. психол.-соц. ин-т, 2003.
- [**Вертгеймер, 1987**] Вертгеймер М. Продуктивное мышление. – М.: Прогресс, 1987.
- [**Воронин и др., 2019**] Воронин А.Н., Гребенщикова Т. А., Кубрак Т. А., Павлова Н. Д. Субъектность сетевого сообщества: сравнение психометрических моделей проявления дискурсивных маркеров в контенте // Вестник Московского государственного областного университета. Серия: Психологические науки. – 2019. – № 3. – Doi: 10.18384/2310-7235-2019-3-6-24.
- [**Григорьев и др., 2022**] Григорьев О.Г., Кузнецова Ю.М., Никитина Е.Н., Смирнов И.В., Чудова Н.В. Каузативно-эмотивный анализ. Ч. I. Методика изучения эмоциональных реакций пользователей социальных сетей // Психологический журнал. – 2022. – Т. 43, № 3.
- [**Загоровская и др., 2016**] Загоровская О.В., Литвинова О.А., Литвинова Т.А. Выявление склонности личности к суицидальному поведению на основе количественного анализа ее речевой продукции // Studia Humanitatis. – 2016. – № 1. – URL: <https://cyberleninka.ru/article/n/vyyavlenie-sklonnosti-lichnosti-k-suitsidalnomu-povedeniyu-na-osnove-kolichestvennogo-analiza-ee-rechevoy-produktsii> (дата обращения: 11.01.2022).
- [**Литвинова и др., 2015**] Литвинова Т.А., Литвинова О.А., Рыжкова Е.С., Бирюкова Е.Д., Середин П.В., Загоровская О.В. Исследование влияния пола и психологических характеристик автора на количественные параметры его текста с использованием программы Linguistic Inquiry and Word Count // Научный диалог. – 2015. – № 12 (48).
- [**Микиртумов и др., 2012**] Микиртумов Б.Е., Ильичев А.Б. Клиническая семантика психопатологии. – СПб.: СПбГПМА, 2007.
- [**Никитина и др., 2022**] Никитина Е.Н., Онипенко Н.К. Семантика и прагматика высказываний с эмотивными предикатами // Сибирский филологический журнал. – 2022. – № 2.
- [**Новиков, 2022**] Новиков Д.А. Интервью директора ИПУ РАН Д.А. Новиковым. – URL: <https://new.ras.ru/mir-nauky/news/vokrug-iskusstvennogo-intellekta-skladyvaetsya-ochen-trevozhnaya-struktura-znaniy-i-kompetentsiy-aka>, (дата обращения: 31.07.2022).
- [**Образцов и др., 2002**] Образцов В.А., Богомолова С.Н. Криминалистическая психология: Методы, рекомендации, практика раскрытия преступлений. – М.: ЮНИТИ, 2002.
- [**Осипов, 2008**] Осипов Г.С., Смирнов И.В., Тихомиров И.А. Реляционно-ситуационный метод поиска и анализа текстов и его приложения // Искусственный интеллект и принятие решений. – М: ЛЕНАНД, 2008. – № 2. – С. 3-10.
- [**Осипов, 2016**] Осипов Г.С., Смирнов И.В. Семантический анализ научных текстов и их больших массивов // Системы высокой доступности. – 2016. – № 1.
- [**Остапенко, 2022**] Остапенко А.Г. и др. Социальные сети и риск-мониторинг / Теория сетевых войн / под ред. Д.А. Новикова. – М.: Телеком. 2022.

- [Поспелов, 1996] Поспелов Д.А. Десять "горячих точек" в исследованиях по искусственному интеллекту // Интеллектуальные системы (МГУ). – 1996. – Т. 1. – Вып. 1-4.
- [Себрант, 2022] Себрант А.Ю. Data Science, черные ящики. – URL: <https://www.youtube.com/watch?v=zvGeLvWZ7yQ&t=257s> (дата обращения: 29.08.2022).
- [Теплов, 1961] Теплов Б.М. Проблемы индивидуальных различий. – М., 1961.
- [Devyatkin et al., 2020] Devyatkin D., Chudova N., Mishlanov V., Salimovsky V., Kadzhaya L. Automatic Identification of Cognitive Actions Constituting: Speech Genres of Scientific Theoretical Text // Proceedings of the Linguistic Forum 2020: Language and Artificial Intelligence LFLAI 2020, Ed.: V. Solovyev, N. Loukachevitch, O. Lyashevskaya, CEUR Workshop Proceedings, 2021.
- [Devyatkin et al., 2021] Devyatkin D., Chudova N., Chuganskaya A., Sharypina D. Methods for Recognition of Frustration-Derived Reactions on Social Media. In: Kovalev S.M., Kuznetsov S.O., Panov A.I. (eds) Artificial Intelligence. RCAI 2021. Lecture Notes in Computer Science. Vol. 12948. – Springer, Cham, 2021. – [https://doi.org/10.1007/978-3-030-86855-0\\_2](https://doi.org/10.1007/978-3-030-86855-0_2).
- [El-Demerdash et al., 2020] El-Demerdash K., El-Khoribi R., Shoman M., & Abdou S. Psychological Human Traits Detection based on Universal Language Modeling // Egyptian Informatics Journal. – 2020. – Doi: 22. 10.1016/j.eij.2020.09.001.
- [Gjurković et al., 2020] Gjurković M., Karan, M., Vukojević, I., Bošnjak M., & Šnajder J. PANDORA Talks: Personality and Demographics on Reddit // Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media. – 2020. – P. 138-152. – Doi: 10.31234/osf.io/94xcp.
- [Moreno et al., 2021] Moreno, J., Martínez-Huertas, J., Olmos, R., Jorge-Botana, G., & Botella, J. Can personality traits be measured analyzing written language? // A meta-analytic study on computational methods. Personality and Individual Differences. – 2021. – 177 (2). – Doi:10.1016/j.paid.2021.110818.
- [Shelmanov et al., 2014] Shelmanov A.O., Smirnov I.V., Methods for Semantic Role Labeling of Russian Texts // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue". – 2014. – Issue 13 (20).
- [Smirnov et al., 2021] Smirnov I., Stankevich M., Kuznetsova Y., Suvorova M., Larionov D., Nikitina E., Savelov M. & Grigoriev O. TITANIS: A Tool for Intelligent Text Analysis in Social Media // Artificial Intelligence 19th Russian Conference, RCAI 2021, Taganrog, Russia, October 11–16, 2021, Proceedings.
- [Tinwala et al., 2021] Tinwala W., & Rauniyar S. Big Five Personality Detection Using Deep Convolutional Neural Networks, 2021. – Doi: 10.20944/preprints202109.0199.v1.

## ФОРМИРОВАНИЕ БАЗ ЗНАНИЙ С УЧЁТОМ ОНТОЛОГИЧЕСКИХ СОГЛАШЕНИЙ НА ПЛАТФОРМЕ IASRAAS\*

Ф.М. Москаленко (*philipmmr@iacp.dvo.ru*)

В.А. Тимченко (*vadima@dvo.ru*)

Институт Автоматики и Процессов Управления  
ДВО РАН, Владивосток

В работе представлены механизмы поддержки задания и проверки онтологических соглашений для орграфовой связанной двухуровневой модели информационных единиц. Применение предложенных подходов в процессе формирования баз знаний должно позволить задавать такие соглашения и осуществлять их проверку без посредников (в лице программистов). Также описывается реализация предложенных решений на платформе IASRAAS, предназначенной для создания и использования облачных систем искусственного интеллекта.

**Ключевые слова:** графовая база данных, база знаний, редактор базы знаний, онтология, онтологические соглашения, IASRAAS.

### Введение

Как отмечается в работе [Грибова и др., 2016], одной из важнейших задач при разработке и сопровождении систем искусственного интеллекта (СИИ) является формирование и поддержка в актуальном состоянии баз данных (БД) и баз знаний (БЗ). Это может выполняться на основе одного из двух подходов: с посредником (в лице разработчика системы) и без. Второй вариант (когда за формирование и наполнение БЗ несет ответственность эксперт, а за структуру – инженер знаний) является более предпочтительным, поэтому в инструментарий для разработки интеллектуальных систем обычно включается подсистема формирования и модификации – редактор БЗ. Примерами последнего являются: Neologism [Cosmin Basca et al., 2008], Moki [Dragoni et al., 2013], Knoodl [Fu, 2015], Ontowiki [Arndt,

---

\* Работа выполнена при частичной финансовой поддержке РФФИ (проект 20-07-00670 и проект 20-01-00449).

2010], SOBOLEO [Braun et al., 2010], PoolParty [Schandl et al. 2010] и WebProtégé [Musen, 2015]. В них делаются попытки дать возможность формировать БЗ и другие сложноструктурированные информационные ресурсы без посредников.

Однако, в основу упомянутых редакторов положена парадигма объектно-ориентированных онтологий [Klepper et al., 2003; Varro et al., 2003; Muller et al., 2005; OMG, 2014], незнакомая большинству реальных и потенциальных экспертов. В связи с этим было предложено формировать графовые базы знаний и данных на основе двухуровневого подхода [Грибова и др., 2015а, 2015б] с использованием специализированного редактора [Грибова и др., 2016]. Их реализация на платформе IASPaas (iaspaas.dvo.ru) [Грибова и др., 2018], предназначенной для создания и использования СИИ, за более чем 10 летний срок применения и постоянной доработки продемонстрировала высокую эффективность и удобство использования. Двухуровневость подхода заключается в том, что вначале, с использованием специализированного метаязыка, формируется оргграф, описывающий метапонятия и их связь (модель онтологии предметной области), который представляет собой язык для описания оргграфов целевой информации (БД или БЗ).

Онтологический подход, являющийся одним из наиболее эффективных способов создания БЗ, подразумевает не только конструирование взаимосвязанных сложноструктурированных информационных единиц, но и должен поддерживать отслеживание выполнения соглашений, заложенных в онтологию инженером знаний, в процессе формирования знаний и информации экспертом. Для решения этой задачи могут использоваться:

- специально разрабатываемые программные продукты;
- функциональные возможности используемого редактора БЗ;
- предоставляемые средой хранения и наполнения БД/БЗ (СУБД) механизмы.

В первом и последнем случаях для решения задачи ввода и проверки онтологических соглашений (ОС) необходимо привлекать программиста, тогда как встраивание таких возможностей в высокоуровневую среду разработки БЗ позволяет, опять же, этого избежать.

В статье описывается модель и методы формирования и слежения за выполнением онтологических соглашений для двухуровневой модели представления информации [Грибова и др., 2015а, 2015б] и модели её редактора [Грибова и др., 2016], а также реализации этих механизмов на платформе IASPaas [Грибова и др., 2018].

## 1. Общие сведения о поддержке онтологических соглашений на платформе IACPaas

В работе [Грибова и др., 2015а] описана орграфовая связанная двухуровневая модель информационных единиц. Для этой модели предложены языки представления орграфов информации и метаинформации. В работе [Грибова и др., 2015b] определены соответствие между этими орграфами и модель порождения орграфов информации по орграфам метаинформации, сохраняющего это соответствие. В работе [Грибова и др., 2016] описывается модель редактора орграфов информации, управляемого орграфами метаинформации, его основная функциональность, а также механизм её расширения.

Можно сказать, что описанная в работе [Грибова и др., 2015b] модель порождения уже поддерживает ряд механизмов для задания и проверки некоторых тривиальных типов онтологических соглашений при формировании информации:

а) используя механизм спецификаторов, можно поставить условие на количество порождаемых от вершины потомков того или иного типа - возможны лишь два варианта: либо в точности 1 потомок (спецификаторы `one` и `сору`) или его возможное отсутствие (спецификаторы `onemm` и `сорумm`), либо множество – непустое (спецификаторы `set`, `list`, `seq`) или возможно пустое (спецификаторы `setmm`, `listmm`, `seqmm`);

б) используя механизм ограничителей, можно ставить условие на значение метки вершины – возможны два варианта: своя метка, метка берётся от другой вершины.

в) используя этот же механизм, можно управлять порождением новой вершины - запрещать или разрешать его (при запрете можно породить лишь дугу, соединяющую с существующей вершиной).

Соответственно и инструментальные средства платформы IACPaas, будучи основанными на данной модели порождения, поддерживают такие возможности «из коробки». Помимо того, что блокируются варианты порождения информационных ресурсов, которые могут приводить к нарушению их соответствия модели онтологии, немаловажной также является поддержка механизма проверки «полноты» – проверка у вершин наличия прямых потомков, порождаемых по метавершинам с нефакультативными спецификаторами у входящих в них дуг.

Вместе с тем, наибольший интерес для инженеров знаний и экспертов представляют сложно-структурированные онтологические соглашения, описывающие закономерности, налагаемые на целые фрагменты знаний предметной области, а не контролирующие лишь указанные три свойства в тех или иных точечных локациях БЗ.



Для поддержки задания и проверки таких онтологических соглашений была усовершенствована двухуровневая модель представления информационных единиц, модель их редактора. Предложенные усовершенствования были реализованы на платформе IASPaas.

1. Разработан язык (онтология) описания онтологических соглашений, их задание в виде логических формул выполняется с использованием двухуровневого подхода – они помещаются в информационные единицы, формируемые по данной онтологии с использованием редактора [Грибова и др., 2016]. На платформе IASPaas язык в виде графового информационного ресурса был помещён в общий Фонд платформы, а сами ОС инженеры знаний задают с использованием редактора информационных ресурсов.

2. Модель редактора онтологий расширена функциями для сопоставления некоторой модели онтологии, для которой разрабатываются онтологические соглашения, с информационной единицей, в которой эти онтологические соглашения хранятся. Соответственно, на платформе IASPaas каждому информационному ресурсу, который представляет модель онтологии, может быть сопоставлен информационный ресурс с онтологическими соглашениями, или такого сопоставления может не быть.

3. Предложен метод проверки сопоставленных онтологий некоторого информационного ресурса онтологических соглашений. Реализация этого метода помещена в программную библиотеку (API) платформы IASPaas, используемую прикладными программистами для разработки СИИ.

4. Модель редактора информационных ресурсов расширена функцией проверки сопоставленных онтологий формируемого информационного ресурса онтологических соглашений. Соответственно и редактор информационных ресурсов платформы IASPaas получил интерфейсные возможности для запуска проверки ОС и отображения её результатов.

## **2. Язык описания онтологических соглашений**

Онтологические соглашения, проверяемые на орграфах информации, описываются на языке логико-математических формул, которые задаются над той моделью онтологии, в соответствии с которой эти орграфы информации формируются. ОС имеет название и представляет собой кванторную формулу. Она состоит из переменной, и формулы, в которую входят термы, в свою очередь, возможно, включающие формулы и термы.

Выделяются следующие типы формул:

- равенство и неравенство термов: «равенство», «неравенство»;
- арифметические отношения над термами: «больше», «больше или равно», «меньше», «меньше или равно»;
- теоретико-множественные отношения над термами: «отношение принадлежности» элемента множеству, «отношение непринадлежности» элемента множеству, множество «есть собственное под-

множество» другого множества, множество «есть несобственное подмножество» другого множества, множество «не есть подмножество» другого множества;

- формулы с пропозициональными связками: «импликация», «равносильность», «конъюнкция», «дизъюнкция», «исключающее или», «отрицание»;
- кванторные формулы: «квантор всеобщности», «квантор существования», «квантор существования и единственности»;
- предикат «прямой предок», моделирующий отношение над терминами, представляющими вершины орграфа информации;
- предикат «соответствие регулярному выражению».

Формально структура ОС в нотации текстового представления, применяемого на платформе [Грибова и др., 2015а, 2015b], выглядит следующим образом:

```
онтологическое соглашение {
  проху кванторные формулы ~ALT{
    ~new квантор всеобщности {
      ~проху переменная и формула {
        ~one ~new -> переменная;
        ~new формула ~ALT{
          ~new равенство {
            ~new левая часть {
              ~new -> терм;
            }
            ~new правая часть {
              ~new -> терм;
            }
          }
          ~new неравенство {
            ~проху -> равенство;
          }
        }
        ~проху арифметические отношения ~ALT{
          ~new больше или равно {
            ~проху два числа {
              ~new левое число {
                ~new -> терм-число;
              }
              ~new правое число {
                ~new -> терм-число;
              }
            }
          }
          ~new больше {
            ~проху -> два числа;
          }
          ~new меньше или равно {
            ~проху -> два числа;
          }
          ~new меньше {
            ~проху -> два числа;
          }
        }
      }
    }
  }
}
```

```

~ргоху теоретико-множественные отношения ~ALT{
  ~new принадлежность элемента множеству {
    ~ргоху терм и множество {
      ~new -> левая часть;
      ~new правая часть
      {
        ~new -> терм-множество;
      } } }
  ~new не принадлежность элемента множеству
  {
    ~ргоху -> терм и множество;
  }
  ~new множество есть собственное подмножество другого
  множества {
    ~ргоху множество и множество {
      ~new левая часть {
        ~new -> терм-множество;
      }
      ~new -> правая часть;
    } }
  ~new множество есть несобственное подмножество
  другого множества {
    ~ргоху -> множество и множество;
  }
  ~new множество не есть подмножество другого
  множества{
    ~ргоху -> множество и множество;
  } }
~ргоху формулы с пропозициональными связками
~ALT{
  ~new импликация {
    ~new условие {
      ~new -> формула;
    }
    ~new заключение {
      ~new -> формула;
    } }
  ~new равносильность {
    ~new левая часть {
      ~new -> формула;
    }
    ~new правая часть {
      ~new -> формула;
    } }
  ~new конъюнкция {
    ~seq конъюнкт
    {
      ~new -> формула;
    }
  }
  ~new дизъюнкция {
    ~seq дизъюнкт {

```

```

    ~new -> формула;
  } }
  ~new исключающее или {
    ~seq -> дизъюнкт;
  }
  ~new отрицание {
    ~new -> формула;
  } }
  ~proху -> кванторные формулы;
  ~new прямой предок {
    ~new предок {
      ~new -> терм;
    }
    ~new потомок {
      ~new -> терм;
    } }
  ~new равенство подграфов {
    ~proху -> равенство;
  }
  ~new соответствие регулярному выражению {
    ~new строка {
      ~new -> терм-строка;
    }
    ~new регулярное выражение [str]
  } } } }
  ~new квантор существования {
    ~proху -> переменная и формула;
  }
  ~new квантор существования и единственности {
    ~proху -> переменная и формула;
  } } }

```

Переменная, входящая в формулу, связана некоторым квантором, имеет имя, а область ее возможных значений задается термом, значением которого является множество вершин орграфа информации. Такое множество может быть задано с помощью термина «множество соответствующих мета-пути вершин» (см. ниже) или с помощью операций над терминами такого типа.

Ниже приводится описание структуры переменной.

```

переменная {
  ~new терм-множество
  ~ALT{
    ~new путь {
      ~new начало пути
      ~ALT{
        ~one ~ref ->  $\Lambda$ ;
        ~one ~ref -> переменная;
      }
      ~сорупп ~new продолжение пути {
        ~seq элемент пути {

```

```

    ~one ~ref ->  $\Lambda$ ;
  } }
  ~coprnm ~new конец пути {
    ~seq -> элемент пути;
  } }
  ~new экстенциональное задание множества {
    ~seq элемент множества {
      ~new -> терм;
    } }
  ~new пересечение множеств {
    ~seq множество {
      ~new -> терм-множество;
    } }
  ~new объединение множеств {
    ~seq -> множество;
  } }
  ~new разность множеств {
    ~new уменьшаемое множество {
      ~new -> терм-множество;
    }
    ~new вычитаемое множество {
      ~new -> терм-множество;
    } }
  ~new квантор интенциональности {
    ~proху -> переменная и формула;
  } }
  ~new квантор объединения {
    ~proху переменная и терм-множество {
      ~one ~new -> переменная;
      ~new -> терм-множество;
    } }
  ~new квантор пересечения {
    ~proху -> переменная и терм-множество;
  } }
  ~new условный терм {
    ~seq альтернатива {
      ~new условие {
        ~new -> формула;
      }
      ~new заключение {
        ~new -> терм;
      } } }
  ~new взять область возможных значений переменной {
    ~one ~ref -> переменная;
  } } }

```

Помимо этого в языке специфицированы следующие типы термов:

- термы, представляющие арифметические выражения: «сумма», «произведение», «разность», «частное», «возведение в степень», «абсолютная величина»;

- термы-кванторные конструкции: «квантор суммирования», «квантор умножения», «квантор конкатенации» «квантор объединения», «квантор пересечения», «квантор преобразования множества», «квантор интенциональности», «йота-оператор»;
- термы, представляющие выражения над множествами: «экстенциональное задание множества», «пересечение множеств», «объединение множеств», «разность множеств», «мощность множества», «наибольший элемент множества», «наименьший элемент множества», «множество соответствующих мета-пути вершин».
- условный терм, имеющий вид  $(f_1 \Rightarrow t_1) \dots (f_n \Rightarrow t_n)$ , где  $f_i$  – формула,  $t_i$  – терм, ( $i = 1, \dots, n$ ),  $n \geq 2$  и моделирующий разветвленный условный оператор;
- терм «вершина-оригинал», значением которого является вершина орграфа информации, зависимым клоном которой является вершина-аргумент;
- терм «конкатенация», представляющий конкатенацию строковых или двоичных констант;
- терм «метка вершины», значением которого является метка вершины орграфа информации;
- термы, представляющие константы различных типов: «пустое множество», «целочисленная константа», «вещественная константа», «логическая константа», «строковая константа», «временная константа», «двоичная константа».
- терм, представляющий собой значение переменной.

Здесь особыми типами термов, характерными именно для применяемой графовой модели информационных ресурсов являются «вершина-оригинал», «метка вершины» (подробнее - в [Грибова и др., 2015a, 2015b]) и «множество соответствующих мета-пути вершин».

Терм «множество соответствующих мета-пути вершин» представляет собой последовательность элементов. Первый элемент последовательности «начало пути» представляет собой либо вершину в орграфе метаинформации, либо определяющее вхождение переменной в кванторную формулу. Каждый последующий элемент последовательности, за исключением, возможно, второго элемента, представляет собой вершину в орграфе метаинформации, являющуюся прямым потомком вершины – предыдущего элемента последовательности. Второй же элемент последовательности может представлять собой вершину, являющуюся как прямым потомком вершины-первого элемента («продолжение пути», в случае, если задается полный путь), так и непрямым потомком вершины-первого элемента («конец пути», в случае, если задается неполный путь). Компонент пути «начало пути» является обязательным для задания, а компонен-

ты «продолжение пути» и «конец пути» могут отсутствовать. Значением термина становятся все вершины целевого информационного ресурса, созданные по указанным в мета-пути понятиям.

Формальное представление термина представлено ниже:

```

терм ~ALT{
  ~one ~ref переменная {
    ~new терм-множество ~ALT{...}
  }
  ~проху термы, связанные с множествами ~ALT{
    ~new -> путь;
    ~new -> экстенциональное задание множества;
    ~проху действия над множествами ~ALT{
      ~new -> объединение множеств;
      ~new -> пересечение множеств;
      ~new -> разность множеств;
      ~new мощность множества {
        ~new -> терм-множество;
      }
      ~new sup {
        ~new -> терм-множество;
      }
      ~new inf {
        ~new -> терм-множество;
      }
    } }
  ~проху арифметические термы ~ALT{
    ~new сумма {
      ~seq слагаемое {
        ~new терм-число ~ALT{
          ~new -> мощность множества;
          ~new -> sup;
          ~new -> inf;
          ~проху -> арифметические термы;
          ~new квантор суммирования {
            ~проху переменная и терм-число {
              ~one ~new -> переменная;
              ~new -> терм-число;
            }
          }
          ~new квантор умножения {
            ~проху -> переменная и терм-число;
          }
          ~new -> условный терм;
          ~new метка вершины {
            ~new -> терм;
          }
          целочисленная константа [int]
          вещественная константа [real]
        }
      }
    }
    ~new произведение {
      ~seq сомножитель {
        ~new -> терм-число;
      }
    }
  }
}

```

```

~new разность {
  ~new уменьшаемое {
    ~new -> терм-число;
  }
  ~new вычитаемое {
    ~new -> терм-число;
  }
}
~new частное {
  ~new делимое {
    ~new -> терм-число;
  }
  ~new делитель {
    ~new -> терм-число;
  }
}
~new возведение в степень {
  ~new основание {
    ~new -> терм-число;
  }
  ~new показатель {
    ~new -> терм-число;
  }
}
~new абсолютная величина {
  ~new -> терм-число;
}
}
~проху кванторные конструкции ~ALT{
  ~new йота-оператор {
    ~проху -> переменная и формула;
  }
  ~new -> квантор интенциональности;
  ~new -> квантор объединения;
  ~new -> квантор пересечения;
  ~new -> квантор суммирования;
  ~new -> квантор умножения;
  ~new квантор конкатенации {
    ~проху переменная и терм-строка {
      ~one ~new -> переменная;
      ~new терм-строка ~ALT{
        ~new -> метка вершины;
        строковая константа [str]
        двоичная константа [blob]
        ~new -> условный терм;
      }
    }
  }
  ~new квантор преобразования множества {
    ~проху переменная и терм {
      ~one ~new -> переменная;
      ~new -> терм;
    }
  }
}
~new -> условный терм;
~new -> метка вершины;
~проху константы ~ALT{
  ~new ["пустое множество"] #с ∅
  -> целочисленная константа;
}

```



```

-> вещественная константа;
логическая константа [bool]
-> строковая константа;
темпоральная константа [date]
-> двоичная константа;
}
~new вершина-оригинал {
  ~one ~ref -> переменная;
}
~new конкатенация {
  ~seq операнд {
    ~new -> терм-строка;
  } } }

```

Описанный язык для задания онтологических соглашений сформирован в виде информационного ресурса типа «метаинформация» на платформе IASaaS её разработчиками и размещён в общем Фонде платформы в разделе Ядро платформы. При необходимости, данный язык может быть ими расширен.

### 3. Формирование онтологических соглашений

С использованием описанного языка формирование онтологических соглашений для некоторой (прикладной) онтологии выполняется в соответствии с орграфовой связанной двухуровневой моделью информационных единиц [Грибова и др., 2015а, 2015б].

На платформе IASaaS для этого используется редактор, являющийся частью инструментария базовой технологии разработки. Он основан на API двухуровневой модели порождения информационных ресурсов. Это, как отмечено в [Грибова и др., 2018], в частности позволяет выполнять данную процедуру без участия программиста, так как применяется стандартный редактор информационных ресурсов, реализованный согласно принципам, изложенным в [Грибова и др., 2016]. Он позволяет работать с информационным ресурсом в форме орграфа, в структурном представлении. Для пользователя, не имеющего навыков программиста, это удобнее текстового представления, которое бы могло применяться в случае задания ОС в виде триггеров и хранимых процедур на некотором языке программирования в реляционных СУБД – при использовании такого подхода к хранению знаний.

Работа инженера на платформе знаний начинается с создания нового целевого информационного ресурса на основе информационного ресурса, представляющего описанный язык (инженеру необходимо единожды получить его в доступ из общего Фонда платформы). Затем этот новый информационный ресурс наполняется экземплярами вершин, описывающих онтологические соглашения (пример см. в разделе 5).

Неотъемлемой частью процесса редактирования является использование прикладной онтологии при задании формул соглашений. Оно состоит в создании ссылок (дуг) на вершины этого информационного ресурса.

Процесс задания онтологических отношений для некоторой модели онтологии завершается связыванием этих информационных ресурсов: прикрепление ресурса с онтологическими соглашениями к информационному ресурсу, хранящему онтологию. Для выполнения прикрепления выполнено расширение модели данных – её части, описывающей атрибуты информационного ресурса. Здесь добавлен новый атрибут, значение которого «указывает» на информационный ресурс с формулами. Причём он может иметь значение (установлена ссылка на ресурс с соглашениями) и не иметь значения (ссылка не установлена, онтология не имеет сопоставленных соглашений).

Помимо расширения самой структуры данных, также был расширен набор методов класса, представляющего информационный ресурс (в рамках модуля, являющегося процессором информационных ресурсов) – для оперирования значением указанного атрибута. Добавлены три метода:

- задать ссылку на информационный ресурс с ОС;
- удалить ссылку на информационный ресурс с ОС;
- получить ссылку на информационный ресурс с ОС.

#### 4. Проверка онтологических соглашений

Для проверки ОС модифицируется алгоритм работы редактора [Грибова и др., 2016]. В него после шага 2.7. добавляются шаги 2.8 и 2.9:

1. автоматическое формирование фрагмента орграфа информации, выполняемое при активации некоторой вершины орграфа – согласно состоянию орграфа информации, содержимому орграфа метайнформации и правилам порождения орграфа информации по орграфу метайнформации выполняются порождения исходящих дуг (и их конечных вершин); если порождаются дуги, исходящие из активной вершины орграфа информации, то их концы становятся активными вершинами (при условии, что автоматическое порождение из них невозможно).

2. выбор пользователем очередного действия и его выполнение – на этом этапе пользователь может:

- 2.1. просмотреть значения полей вершины орграфа информации;
- 2.2. активировать вершину орграфа информации;
- 2.3. деактивировать активную вершину в орграфе информации;
- 2.4. у активной вершины орграфа информации:

- инициировать изменение значений ее полей, с их последующим изменением и сохранением,  
- просмотреть исходящие из неё дуги,  
- инициировать формирование исходящих из неё дуг, с их последующим созданием (включая их конечные вершины, которые становятся активными),  
- сделать её текущей, т.е. начальной вершиной подграфа орграфа информации для последующего редактирования этого подграфа (кроме терминальных вершин),

- изменить порядок исходящих из нее дуг,
  - удалить входящую в нее дугу (кроме текущей вершины);
- 2.5. вернуться к предыдущему состоянию редактирования подграфа, начальная вершина которого является текущей;
- 2.6. отобразить в интерфейсе редактора весь порожденный подграф орграфа информации, начальная вершина которого является текущей (при этом все вершины этого подграфа становятся активными);
- 2.7. скрыть в интерфейсе редактора весь порожденный подграф орграфа информации, начальная вершина которого является текущей (при этом все вершины этого подграфа, кроме начальной, становятся неактивными).
- 2.8. инициировать проверку ОС, если такая проверка ещё не ведётся;
- 2.9. отобразить результаты проверки ОС, если они готовы, или отобразить сообщение о выполнении проверки.

**Алгоритм метода проверки множества ОС, сопоставленных модели онтологии, состоит в следующем:**

1. если к онтологии (метаинформации) текущего информационного ресурса, не прикреплен ресурс с онтологическими соглашениями, то выдать сообщение об этом и завершить работу;
2. выполнить цикл по онтологическим соглашениям, которые находятся в информационном ресурсе, прикрепленном к онтологии (метаинформации) текущего информационного ресурса (при наличии):
  - 2.1. вызвать метод для проверки истинности формулы, находящейся во главе онтологического соглашения;
  - 2.2. если проверка формулы дала результат «ложь» - выдать сообщение о нарушении текущего онтологического соглашения и завершить работу, иначе - перейти к следующему онтологическому соглашению;
  3. выдать сообщение о выполнении всех онтологических соглашений.

Для обработки всех типов формулы разрабатывается отдельный под-метод, вычисляющий тип формулы и её истинность (вызывая под-метод вычисления значения термина для всех входящих в формулу термов) и возвращающий соответствующее логическое значение. Таким же образом для вычисления значения термина разрабатывается отдельный под-метод, вычисляющий его тип и значение. В общем случае схема работы этих под-методов представляет собой оператор SWITCH-CASE, где в условии SWITCH вычисляется тип формулы или термина, а в ветках CASE выполняется вычисление значения.

Реализация указанных функциональных возможностей на платформе IASaaS начинается с расширения её процессора информационных ресурсов, а именно – снабжение его реализацией метода проверки онтологических соглашений. Вызов данного метода помещается не только в редактор (см. далее) разработчиками платформы, но и сделан доступным в API разработчика, то есть становится доступен прикладным разработчикам – для вызова в рамках кода блоков продукций их агентов.

Модификация редактора информационных ресурсов состояла в следующем. В его интерфейс добавлена кнопка инициации проверки онтологических соглашений. Она отображается лишь в том случае, когда в редакторе выполняется формирование целевого ресурса, к онтологии которого подключен ресурс с онтологическими соглашениями (в противном случае отображается сообщение об отсутствии такого ресурса). Нажатие на данную кнопку приводит к:

а) инициации проверки онтологических соглашений отдельным агентом, вызываемым редактором, который в свою очередь вызовет описанный метод API для проверки ОС,

б) отображению сообщения о том, что проверка инициирована и кнопкой проверки наличия результата.

При нажатии на последнюю, в случае отсутствия результата проверки, которая может занимать продолжительное время при работе с базами знаний, включающими несколько десятков/сотен тысяч вершин, отображается прежний интерфейс, а при его готовности (наличии) – выводится итоговое сообщение вида «отсутствие нарушений» или «нарушено онтологическое соглашение ...».

## 5. Пример использования

Для демонстрации описанных наработок приведем пример задания и проверки некоторых тривиальных онтологических соглашений из области медицинской диагностики. Для решения такой задачи обычно строится набор баз знаний, основой которого является база, где описываются все возможные наблюдения. Для простоты описания условимся, что признаки, описывающие пациентов могут иметь лишь качественные значения. Рассматриваемые онтологические соглашения – следующие:

а) все значения каждого признака попарно различны,

б) каждый признак имеет более 1 значения.

Тогда, используя метаязык, формируется модель онтологии, задаваемая следующим образом:

```
Структура базы наблюдений {  
  ~setmm наблюдение {  
    ~set значение [str]  
  }  
}
```

Далее при формировании целевых информационных баз знаний о наблюдениях экспертами благодаря заложенному в платформу принципу двухуровневого формирования информационных единиц и механизму его поддержки при вызове метода создания очередного потомка у каждой вершины с метапонятием «значение» будет отслеживаться ОС (а) и в случае задания значения, совпадающего с уже существующим, вместо созда-

ния новой вершины с указанной меткой и дугой к ней метод вернёт ошибку (которая будет отображена в Редакторе – при редактировании БЗ, или должна быть обработана пользовательским кодом – в рамках кода блока продукций агента прикладного сервиса).

Второе ОС не может быть проверенным упомянутым выше базовым механизмом платформы, так как ограничения, налагаемые на число томов вершины, доступные в этом механизме к использованию, допускают указание на необходимость создания 0, 1 или многих вершин. Поэтому для проверки ОС (б) используется описанный в статье механизм:

а) формируется отдельный информационный ресурс, в котором указывается данное ОС:

```
ОС для баз наблюдений <<< Язык описания ОС {
  соглашение [признак имеет два или более значения] {
    квантор всеобщности {
      переменная [признак] {
        терм-множество {
          множество соответствующих мета-пути вершин {
            начало пути {
              вершина {
                Л [-> Структура наблюдений$/признак;]
              }
            }
          }
        }
      }
    }
  }
  формула {
    больше или равно {
      левое число {
        терм-число {
          мощность множества {
            терм-множество {
              множество соответствующих мета-пути вершин {
                начало пути {
                  переменная [-> ОС для баз наблюдений$
                    /признак имеет два или более значения
                    /квантор всеобщности/признак;]
                }
              }
            }
          }
        }
      }
      конец пути {
        элемент пути [1] {
          Л [ -> Структура наблюдений$
            /признак/значение;]
        }
      }
    }
  }
  правое число {
    терм-число {
      целочисленная константа [2]
    }
  }
}
```

В графическом виде данный информационный ресурс представлен на рис. 1.

Затем полученный ресурс «подключается» в качестве источника онтологических соглашений к «Структура наблюдений» (в меню «контекст» Редактора информационных ресурсов – рис. 2).



Рис. 1. Редактор с информационным ресурсом «ОС для баз наблюдений»

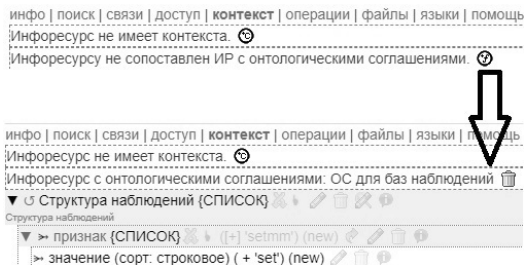


Рис. 2. Подключение в Редакторе информационного ресурса «Структура наблюдений» ресурса «ОС для баз наблюдений» в качестве хранилища онтологических соглашений

После этого все эксперты и прочие пользователи, формирующие базы знаний, созданные по «Структура наблюдений» имеют возможность выполнить проверку данного онтологического соглашения (и прочих, при их наличии в «ОС для баз наблюдений» – рис. 3.



*Рис. 3. Этапы проверки онтологического соглашения в Редакторе информационного ресурса «База наблюдений»*

Здесь вначале необходимо в меню «инфо» запустить проверку, затем периодически обновлять страницу, нажимая на кнопку-иконку «лупа». По окончании проверки будет получен один из 2 результатов: успешное выполнение онтологических соглашений или провал при проверке некоторого онтологического соглашения.

## Заключение

В работе представлен язык описания онтологических соглашений для орграфовой связной двухуровневой модели информационных единиц. С его использованием описываются всевозможные формулы, составляющие ОС, в виде их суперпозиции. Инженер знаний при помощи редактора может сформировать требуемые соглашения для определённой модели онтологии и сопоставить их ей. Предложены методы реализации функциональных возможностей по проверке таких онтологических соглашений, выполняемой в процессе формирования экспертами своих информационных ресурсов, и информированию пользователя о её результатах.

Полученные результаты реализованы на облачной платформе IASaaS, предназначенной для разработки и использования облачных интеллектуальных сервисов. В частности, был доработан входящий в состав инструментария платформы редактор. Внесённые изменения были успешно опробованы при формировании ряда медицинских БЗ.

Вместе с тем, мы видим и несколько путей дальнейшего развития редактора и применяемого подхода. Например, в текущей реализации сложно по отображаемым результатам проверки (в случае невыполнения какого-то ОС) понять лишь по имени соглашения, что и где необходимо модифицировать в БЗ, чтобы данное ОС не было нарушено. Необходим механизм предложения более детального информирования, а также быстрого перехода к месту в графе, где выявлено нарушение соглашения.

Решением первой задачи может быть расширение языка описания соглашений, которое позволит инженеру знаний задавать вспомогательный текст с описанием возможных изменений в ресурсе с целью приведения его в соответствие с ОС.

Решение второй задачи может быть получено одним из двух путей:

а) модификация метода, вычисляющего значение формулы, а именно - преобразование результата его работы из просто логического значения («выполнено» или «не выполнено»), в некий комплексный результат (где также присутствует индикация места в графовой БЗ, где формула «не выполнена»);

б) формирование (хотя бы некоторых, при возможности) ОС не как «глобальных» для всей БЗ, а с привязкой к использованным в ОС вершинам из метаинформации (это потребует расширение применяемой на платформе IASPaas модели данных и ляжет дополнительной работой на инженера знаний), что позволит при создании и модификации вершин, созданных по этим метавершинам, сразу иметь и проверяемое ОС и место, где происходит редактирование, нарушающее его.

Интересным направлением дальнейших работ является реализация «самоприменения» описанного механизма ОС, а именно – задание ОС для самого языка описания ОС. К таким ОС можно отнести, например, следующее: терм, представляющий собой область возможных значений определяемой переменной, не может содержать вхождений этой же самой переменной.

Помимо поддержки внесения и проверки онтологических соглашений немаловажной составляющей средств (инструментов) формирования БЗ является поддержка контекстно-зависимого редактирования. В настоящее время на платформе IASPaas уже имеются частично отвечающие за это механизмы «Присваивание» и «Выбор» (см. [Грибова и др., 2016]), но планируется и дополнительное развитие инструментальных средств платформы.

### Список литературы

- [Грибова и др., 2015a] Грибова В.В., Клещев А.С., Москаленко Ф.М., Тимченко В.А. Двухуровневая модель сложноструктурированных информационных единиц, соответствующая метафоре анкетирования // Научно-техническая информация. Сер. 2. – 2015. – № 10. – С. 1-10.
- [Грибова и др., 2015b] Грибова В.В., Клещев А.С., Москаленко Ф.М., Тимченко В.А. Модель порождения орграфов информации по орграфу метаинформации для двухуровневой модели сложноструктурированных информационных единиц // Научно-техническая информация. Сер. 2. – 2015. – № 12. – С. 26-38.



- [Грибова и др., 2016] Грибова В.В., Клещев А.С., Москаленко Ф.М., Тимченко В.А. Модель редактора сложноструктурированных информационных единиц, управляемого метаинформацией, и его реализация // Научно-техническая информация. Сер. 2. – 2016. – № 2. – С. 1-13.
- [Грибова и др., 2018] Грибова В.В., Клещев А.С., Москаленко Ф.М., Тимченко В.А., Федорищев Л.А., Шалфеева Е.А. Облачная платформа IACРaaS для разработки оболочек интеллектуальных сервисов: состояние и перспективы развития // Программные продукты и системы. – 2018. – Т. 31, № 3. – С. 521-526. – Doi: 10.15827/0236-235X.031.3.527-536.
- [Arndt, 2010] Natanael Arndt, OntoWiki: a tool providing support for agile, distributed knowledge engineering scenarios. – URL: <http://aksw.org/Projects/OntoWiki.html> (дата обращения: 10.09.2022).
- [Braun et al., 2010] Simone Braun and Valentin Zacharias, SOBOLEO – Editor and Repository for Living Ontologies. – 2010. – URL: <http://knowledge-maturing.com/services-and-tools/tools/> (дата обращения: 10.09.2022).
- [Cosmin Basca et al., 2008] Cosmin Basca et al., Neologism: Easy Vocabulary Publishing. 2008 – URL: <http://ceur-ws.org/Vol-368/paper10.pdf> (дата обращения: 10.09.2022).
- [Dragoni et al., 2013] Dragoni M., Ghidini C., Bosca A. Multilingual MoKi: How to Manage Multilingual Ontologies in a Wiki / ed. Philipp Cimiano, Miriam Fernández, Vanessa Lopez, Stefan Schlobach, Johanna Völker // ESWC (Satellite Events). – Springer, 2013. – P. 162-166.
- [Fu, 2015] Guoxing Fu, Weinian Rao. An introduction of Knoodl. – URL: <http://tinman.cs.gsu.edu/~raj/8711/sp11/presentations/knoodlReport.pdf> (дата обращения: 10.07.2015).
- [Kleppe et al., 2003] Kleppe A., Warmer S., Bast W. MDA Explained. The Model Driven Architecture: Practice and Promise. – Addison-Wesley, 2003. – 192 p.
- [Muller et al., 2005] Muller P.-A., Fleurey F., Jezequel J.-M. Weaving Executability into Object-Oriented Metalanguages / ed. L. S.K. Briand // MoDELS 2005. LNCS. – Springer, 2005. – Vol. 3713. – P. 264-278.
- [Musen, 2015] Musen, M.A.: The Protégé project: A look back and a look forward // AI Matters. – 2015. – 1 (4). – P. 4-12.
- [OMG, 2014] Object Management Group. Model Driven Architecture. – URL: <http://www.omg.org/mda> (дата обращения: 10.09.2022).
- [Schandl et al. 2010] Thomas Schandl and Andreas Blumauer. PoolParty: SKOS Thesaurus Management Utilizing Linked Data. – 2010 – URL: [https://www.researchgate.net/profile/Andreas-Blumauer/publication/220853892\\_PoolParty\\_SKOS\\_Thesaurus\\_Management\\_Utilizing\\_Linked\\_Data/links/00b4952ce82c414b71000000/PoolParty-SKOS-Thesaurus-Management-Utilizing-Linked-Data.pdf](https://www.researchgate.net/profile/Andreas-Blumauer/publication/220853892_PoolParty_SKOS_Thesaurus_Management_Utilizing_Linked_Data/links/00b4952ce82c414b71000000/PoolParty-SKOS-Thesaurus-Management-Utilizing-Linked-Data.pdf) (дата обращения: 10.09.2022).
- [Varro et al., 2003] Varro D., Pataricza A. VPM: A visual, precise and multilevel metamodeling framework for describing mathematical domains and UML // Journal of Software and Systems Modeling. – 2003. – Vol. 2 (3). – P. 187-210.

УДК 004.82

## ОБРАЗОВАТЕЛЬНАЯ ОНТОЛОГИЯ ONTOMATH<sup>EDU</sup>: ПРОБЛЕМЫ ОНТОЛОГИЧЕСКОГО ИНЖИНИРИНГА\*

О.А. Невзорова (*onevzoro@gmail.com*)

М.В. Фалилеева (*mmwff@mail.ru*)

Е.К. Липачёв (*elipachev@gmail.com*)

Л.Р. Шакирова (*liliana008@mail.ru*)

А.Э. Дюпина (*anastasiya.dupina@yandex.ru*)

Казанский федеральный университет, Казань

А.В. Кириллович (*alik.kirillovich@gmail.com*)

Межведомственный суперкомпьютерный центр РАН, Казань

Статья посвящена задачам онтологического инжиниринга в области школьного образования. На основе предложенного подхода к проектированию онтологии в области школьной математики разработана образовательная полилингвальная математическая онтология OntoMath<sup>Edu</sup> для школьного курса планиметрии. Описаны особенности организации образовательных онтологий, отражающих специфику предметного обучения в различных странах, в том числе учет методики преподавания и языка обучения. Специфические особенности организации предметной области позволяют учитывать, в том числе, и уровень подготовки обучающихся. Для решения этих задач предложена новая структура организации онтологий с образовательными проекциями, образовательными уровнями и пререквизитами.

**Ключевые слова:** онтологический инжиниринг, представление математического знания, математическое образование, образовательная проекция, пререквизиты, OntoMath<sup>Edu</sup>.

### Введение

Онтологический инжиниринг в области математического образования – достаточно новая задача, требующая учета при разработке онтологии как специфики предметной области, так и требований соответствующей

---

\* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 19-29-14084).

образовательной системы. В настоящее время онтологии достаточно широко применяются в области образования и моделируют различные аспекты образовательного процесса. Так, можно указать следующие области применения онтологий в образовании: обучающие системы, построение учебных планов, формирование компетенций, управление образовательным процессом, виртуальные обучающие среды, динамика (индивидуализация) процессов обучения. Наиболее трудоемким, безусловно, является проектирование предметных онтологий, применяемым для поддержки функционирования обучающих систем и предметных курсов.

Одной из предметных онтологий в области школьной математики является новая онтология *OntoMath<sup>Edu</sup>* (<https://github.com/CLLKazan/OntoMathEdu>). Эта онтология предназначена для встраивания в школьные математические курсы, а также организации учебного материала с использованием современных семантических технологий.

Одной из важных задач при проектировании онтологий является сопряжение онтологической и лингвистической моделей представления знаний. Онтологическая модель задает способ структурирования знаний, лингвистическая модель определяет способы функционирования языковых единиц в текстах. Другими словами, лингвистическая модель определяет лексические и синтаксические модели концептуальных (онтологических) единиц. Разработанная онтология *OntoMath<sup>Edu</sup>* относится к классу лингвистических (терминологических) онтологий, которая фактически представляет собой базу знаний онтологического типа о понятийной системе и лексико-терминологическом составе предметной области, поэтому данная модель успешно применяется для автоматической обработки математических текстов.

Онтология *OntoMath<sup>Edu</sup>* является центральным компонентом разрабатываемой математической цифровой образовательной платформы, которая предназначена для решения таких задач как: (1) автоматическое тестирование знаний; (2) автоматическая рекомендация образовательных материалов в соответствии с индивидуальным учебным планом; (3) семантическое аннотирование учебных материалов.

В текущем состоянии онтология *OntoMath<sup>Edu</sup>* детально описывает предметную область планиметрии и используется для встраивания в курс планиметрии для общеобразовательной средней школы, разрабатываемый в Казанском федеральном университете. Поскольку онтология *OntoMath<sup>Edu</sup>* ориентирована на образовательные приложения, то важными компонентами онтологии является дидактические отношения, которые образуют независимое измерение онтологии и выступают в качестве аналога отношения предварительных требований (отношение *prerequisite*).

Разработанная математическая онтология как предметная онтология представляет самостоятельный интерес с точки зрения классификации знаний выделенной предметной области. Это обстоятельство позволяет сформулировать другую важную цель, а именно представление разработанной модели онтологии в качестве ядра облака Открытых связанных данных для математического образования. Онтология OntoMath<sup>EdU</sup> была представлена на конференциях SICM [Kirillovich et al, 2020a; Kirillovich et al, 2020b]. В настоящей статье мы описываем процесс онтологического инжиниринга, целью которого является создание образовательной математической полилингвальной онтологии, основные проблемы проектирования и подходы к их решению.

## 1. Онтологические модели для задач образования

Интеллектуальные программные инструменты, как правило, используют базы знаний онтологического типа. В инженерии знаний термин «онтология» определяется как концептуальная модель предметной области, представленная на формальном языке, позволяющем использовать компьютерные инструменты обработки информации (см., например, [Guarino et al, 2009]). В настоящее время для формализованного представления математического знания разработан ряд тезаурусов и онтологий. Отдельно, отметим онтологию OntoMath<sup>PRO</sup>, разработанную в Казанском университете [Nevzorova et al, 2014]. Эта онтология содержит определения, как общепринятых математических понятий, так и развивающуюся терминологию профессиональной математики. На основе онтологии OntoMath<sup>PRO</sup> разработаны сервисы поиска по математическим текстам и рекомендательная система для коллекций физико-математических документов [Elizarov et al, 2016]. Семантические модели активно внедряются в практику электронного образования для моделирования индивидуальной траектории обучения, семантического аннотирования образовательных материалов, оценки знаний, а также онтологического моделирования когнитивного профиля обучаемого [Paquette et al, 2021; Муромцев, 2020].

Можно отметить растущую практику использования открытых связанных данных в образовании [D'Aquin, 2016]. В образовательном контексте связанные данные используются для решения проблем функциональной совместимости образовательных данных и ресурсов, для обогащения образовательного контента, а также для персональных рекомендаций образовательного контента [Pereira et al, 2018].

Онтологический подход в последние годы нашел применение и в обучении школьной математике [Varana et al, 2018]. Экспериментальное исследование авторов [Tzoumpa et al, 2020] послужило основанием для утверждения, что онтологический подход посредством абстрактного и комбинированного мышления помогает учащимся лучше понимать геометрические понятия и их динамические взаимосвязи.

При разработке образовательных программ обязательным этапом является определение зависимостей учебных модулей и последовательности их изучения [Муромцев, 2020]. Семантические зависимости между учебными модулями задаются с помощью пререквизитов рассматриваемой учебной дисциплины. В образовательных онтологиях дидактическое отношение *пререквизит* связывает концепты онтологии на основе стандартов обучения соответствующей системы образования [Kirillovich et al, 2020c].

## 2. Проектирование полилингвальной математической онтологии OntoMathEdu

Подробно опишем основные проблемы и методы их решения при проектировании полилингвальной математической онтологии, предназначенной для учебного курса планиметрии средней школы.

Построение полилингвальной образовательной онтологической модели предметной области планиметрии потребовало формализации понятий содержания курса планиметрии, отношений между ними и построения полноценных альтернативных языковых систем с учетом организации учебного процесса. Для этого требовалось решить следующие задачи:

- сформировать базу учебных материалов на различных языках;
- извлечь геометрические понятия из учебных материалов;
- установить соответствия терминов онтологии на разных языках в различных системах обучения;
- построить иерархии геометрических понятий;
- построить отношения между понятиями предметной области;
- ввести в онтологию специфические отношения, отражающие процесс обучения выбранному предмету школьной математики;
- описать понятия и все его отношения (определения, изображения, ссылки на внешние ресурсы и др.).

В качестве первого курса школьной математики при проектировании онтологии OntoMath<sup>Edu</sup> был выбран курс планиметрии, изучаемый в школах Российской Федерации в 7-х–9-х классах, причем в Республике Татарстан данный курс ведется в ряде школ на татарском языке. В Великобритании аналогичный курс рассчитан на учеников средних школ (Secondary Education).

Разработка онтологии OntoMath<sup>Edu</sup> преследует цель создания полилингвальной онтологии с терминологическими системами математических понятий на русском, английском и татарском языках. Терминологическая база онтологии для курса планиметрии на русском и татарском языках базируется на методических системах обучения планиметрии, принятых в Российской Федерации. При этом татарская терминология в онтологии

представляет собой, в основном, переведенные эквиваленты терминов на русском языке. Отношения между русской и английской терминологическими частями онтологии потребовали разработки отдельной модели их взаимосвязей. Предложенные решения по построению такой модели взаимосвязей позволяют в дальнейшем создавать и другие полилингвальные образовательные предметные онтологии.

**Отбор понятий.** При отборе геометрических понятий использовались экспертные знания авторов статьи, а также учебные и справочные материалы на русском, английском и татарском языках. Можно указать учебник по геометрии Л.С. Атанасяна и др., дополнительные учебные пособия по элементарной геометрии Г.К. Гордина и др.; учебник геометрии Л.С. Атанасяна на татарском языке (пер. с рус. З.Х. Билалова, В.З. Закиров), русско-татарские и русско-татарско-английские толковые словари и справочники математических терминов (И.Г. Галютдинова, Л.И. Галиевой, Л.Л. Салеховой и др), русско-татарский и татарско-русский словари Lingvo 12.0 и др.; учебники по геометрии на английском языке, такие как Africk Henry (2013), Alexander, D.C., Koeberlein, G.M. (2016), Ameen, Ayesha&Khan, Khaleel&PadmajaRani, B. (2012), Cummins, J., Carter, J.A., Cuevas, G.J., Day, R, Malloy, C. (2012) и др., русско-английские словари «Мультитран», Lingvo 12.0 и др.

Выбор языка и системы обучения планиметрии потребовало учета в модели онтологии следующих факторов:

- методические особенности процесса обучения планиметрии в системах образования России и Великобритании, а именно наличие геометрических понятий в курсе планиметрии для школьников; последовательность изучения понятий и устанавливаемые взаимосвязи между ними, предлагаемые определения геометрических понятий и др.;
- образовательная направленность онтологии требует выделения уровней подготовки учащихся по предмету (на каждой образовательной ступени изучения планиметрии предлагается свой набор геометрических понятий и связей между ними);
- лингвистические особенности русского и английского языков при переводе математической терминологии, в том числе, выбор используемого варианта перевода термина при наличии нескольких вариантов (например, высота/height, altitude; круг/circle, disk) или проблема отсутствия термина в школьной программе (kite/дельтоид, complementary angle /угол, дополняющий до угла 90 градусов) и др.

**Описание понятий.** Важной задачей при построении описания понятия в онтологии является отбор определений данного понятия из различных источников. В существующих школьных учебниках нередко приво-

дятся различные определения отдельных геометрических понятий. Например, приведем определения понятия *многоугольник* в основных учебниках по геометрии на русском языке:

- 1) простая замкнутая ломаная (Погорелов А.В.);
- 2) плоский многоугольник или многоугольная область – конечная часть плоскости, ограниченная многоугольником (Погорелов А.В.);
- 3) фигура, составленная из отрезков АВ, ВС, ..., FA, так, что смежные отрезки не лежат на одной прямой, несмежные отрезки не имеют общих точек (Атанасян Л.С.);
- 4) фигура, состоящая из сторон многоугольника и его внутренней области (Атанасян Л.С.);
- 5) замкнутая ломаная, не имеющая самопересечений, ограничивает многоугольник (Шарыгин И.Ф.);
- 6) фигура, образованная простой замкнутой ломаной и ограниченная ею часть плоскости (Смирнова И.М.).

Приведенные определения относят концепт *многоугольник* в принципиально различные классы (многоугольник как часть плоскости, или многоугольник как ломаная, или совокупность ломаной и части плоскости). Выбор определения позволяет установить положение геометрического понятия в таксономии и в дальнейшем построить многочисленные отношения между рассматриваемым понятием и другими понятиями (например, площадь и периметр, принадлежность/непринадлежность точек многоугольнику и др.). В частности, для многоугольника наш выбор связан с определением, связывающим многоугольник с частью плоскости, поскольку нельзя измерить площадь ломаной.

**Проекция онтологии.** Для выделения системы терминов, определяющих область понятий курса планиметрии той или иной системы образования, были созданы проекции онтологии  $\text{OntoMath}^{\text{Edu}}$  (далее – проекция онтологии). Проекция онтологии – это система концептов онтологии и связывающих их отношений, являющаяся формализованной моделью содержания школьной математики определенной системы образования.

В онтологии  $\text{OntoMath}^{\text{Edu}}$  представлены две проекции онтологии: на русском языке в системе образования РФ, вторая на английском языке в системе образования Великобритании. Проекция онтологии связаны друг с другом через единые концепты геометрических понятий, присутствующих одновременно в двух системах обучения планиметрии. Каждый концепт онтологии имеет название на русском и английском языках. В проекции онтологии дополнительно размечаются образовательные уровни. Образовательный уровень онтологии  $\text{OntoMath}^{\text{Edu}}$  включает совокупность концептов предметной области, изучаемых на определенной ступени обу-

чения в соответствии с данной образовательной системой. Для разметки принадлежности концепта онтологии образовательному уровню введено свойство «Образовательный уровень».

В проекции онтологии на русском языке выделяются следующие значения свойства «Образовательный уровень»: 7 класс, 8 класс, 9 класс, 8 профильный класс, 9 профильный класс, дополнительный уровень. В проекции онтологии на английском языке значения Key Stage 1, Key Stage 2, Key Stage 3 и Key Stage 4.

Каждый последующий образовательный уровень включает в себя множества концептов нижерасположенных уровней.

Так, объединение концептов со свойствами образовательный уровень 7 класс, 8 класс и 8 профильный класс позволяет выделить концепты, соответствующие уровню подготовки учащегося закончившему обучение в 8 профильном математическом классе. Таким образом, выделенные образовательные уровни моделируются по отношению включения.

Выделение проекций и образовательных уровней в онтологии является новым шагом в проектировании полилингвальных образовательных 8.

### 3. Структура Онтологии $\text{OntoMath}^{\text{Edu}}$

На основе предложенного подхода к проектированию онтологии в области школьной математики разработана образовательная полилингвальная математическая онтология  $\text{OntoMath}^{\text{Edu}}$  для школьного курса планиметрии.

Формально, онтология  $\text{OntoMath}^{\text{Edu}}$  представляется в виде кортежа  $O = \langle C, Cr, Rel, Hc, Hrel, A \rangle$ , где  $C$  – множество классов,  $Cr$  – множество реализованных отношений (подкласс  $C$ ),  $Rel \subseteq C \times C$  – множество свойств классов,  $Hc: C \times C$  – иерархия классов,  $Hrel: R \times R$  – иерархия отношений,  $A$  – множество аксиом над классами и ограничений отношений.

Проекция онтологии на образовательную систему

$$O^* = \langle C^*, Rel^*, Hc^*, Hrel, A^* \rangle,$$

где  $C$  – множество классов,  $Cr$  – множество реализованных отношений (подкласс  $C$ ),  $Rel \subseteq C \times C$  – множество свойств классов,  $Hc: C \times C$  – иерархия классов,  $Hrel: R \times R$  – иерархия отношений,  $A$  – множество аксиом над классами и ограничений отношений.

В структуре онтологии  $\text{OntoMath}^{\text{Edu}}$  выделяются следующие слои представления, далее называемые уровнями: лингвистический уровень, уровень предметной онтологии, уровень верхней онтологии.

Уровень онтологии верхнего уровня содержит в качестве модели представления онтологию UFO (Unified Foundational Ontology) [Guizzardi, 2005].



Уровень предметной онтологии содержит концепты предметной области школьной математики. Концепты связаны с внешними ресурсами из облака Открытых связанных данных (Linked Open Data, LOD), такими как проект DBpedia (<https://www.dbpedia.org>), онтологии ScienceWISE и OntoMath<sup>PRO</sup>. Кроме того, на основе схемы именования ресурсов MMT (Module system for Mathematical Theories) [Müller et al, 2017], концепты могут быть связаны с онтологией MitM [Dehaye et al, 2016], и, через нее, с концептами различных систем компьютерной алгебры.

Уровень предметной онтологии разработан в соответствии со следующими принципами моделирования:

1. Концептуальная модель школьной математики. OntoMath<sup>Edu</sup> отражает набор концептов школьной математики, который применяется в учебной математической литературе.
2. Строгое соблюдение онтологических различий, определяемых в онтологии верхнего уровня. Например, мы явно разделяем концепты на типы и на роли.
3. Материализация предметных отношений. Многоместные математические отношения в онтологии представлены как концепты, а не как объектные свойства. Благодаря этому математические отношения являются сущностями первого порядка и могут выступать в качестве объектов утверждений.
4. Многоязычность. Концепты онтологии содержат названия на английском, русском и татарском языках.

В структуре онтологии выделяются: иерархия объектов; иерархия материализованных отношений; сеть точек зрения. Текущая версия OntoMath<sup>Edu</sup> содержит 916 концептов, при этом иерархия объектов включает 675 классов, иерархия материализованных отношений 212 классов. В онтологии введено 1280 отношений «класс-подкласс».

Концепты-классы имеют названия на английском, русском и татарском языках. В соответствии с требованиями онтологии верхнего уровня объекты предметной онтологии разделяются на концепты-типы и концепты-роли.

Верхний уровень иерархии типов включает концепты *Фигура на плоскости*, *Аксиома планиметрии*, *Теорема планиметрии*, *Задача планиметрии*, *Единица измерения*, *Инструмент измерения или построения геометрических фигур*.

Онтология определяет следующие отношения, представленные объектными и аннотационными свойствами, и их под свойствами: (1) отношение «иметь аргумент»; (2) отношение онтологической зависимости; (3) отношение «Часть-целое»; (4) отношение предметности; (5) отношение «Пререквизит» и отношение «Образовательный уровень»; (6) отношение «Внешний ресурс».

Онтология содержит иерархию материализованных отношений. Материализованные отношения служат средством для представления утверждений формальной математической теории, основанной на логике предикатов первого порядка, в виде RDF-графа.

Утверждение формальной теории вида  $R(c_1, \dots, c_n)$  представляется в терминах онтологии в виде общей схемы RDF-триплетов (рис. 1):

$\_:\text{rel rdf:type map}(R).$ $\_:\text{rel ome:hasArgument map}(c_1).$ $\dots$ $\_:\text{rel ome:hasArgument map}(c_n).$	$\text{map}(R)$ – отображение имени предиката $R$ на URI материализованного отношения из $\text{OntoMath}^{\text{Edu}}$  $\text{map}(c_i)$ – отображения имен констант-аргументов предиката на URI объектов из $\text{OntoMath}^{\text{Edu}}$
---	---

Рис. 1. Модель материализованного отношения в  $\text{OntoMath}^{\text{Edu}}$

Таким образом, отношения между концептами являются сущностями первого порядка, и могут выступать в качестве объекта утверждений. Примером утверждения вида  $R(c_1, \dots, c_n)$  может быть утверждение  $\text{tangent}(a,b,c)$ , т.е. «прямая  $a$  касается окружности  $b$  в точке  $c$ ».

Верхний уровень иерархии материализованных отношений содержит концепты-классы *Взаимное расположение геометрических фигур на плоскости, Отношение сравнения геометрических фигур, Преобразования плоскости, Метрическое свойство геометрической фигуры*.

**Сеть точек зрения.** Помимо универсальных утверждений о математических понятиях, онтология содержит утверждения, привязанные к отдельным точкам зрения. Точки зрения моделируются с использованием шаблона проектирования «Descriptions and Situations» и базируются на онтологии верхнего уровня DOLCE+DnS Ultralite [Borgo et al, 2010].

В настоящее время в онтологии введены следующие виды точек зрения:

1. Определения. Один и тот же концепт может иметь разные определения в соответствии с разными точками зрения.
2. Образовательные уровни. В онтологии имеются различные образовательные уровни, которые относятся к различным образовательным системам.

**Внешние ссылки.** Онтология  $\text{OntoMath}^{\text{Edu}}$  связана в настоящее время с ресурсом DBpedia. Полуавтоматически построены и проверены вручную 142 связи с данным ресурсом на основе подхода, предложенного в [Kirillovich et al, 2020b].

**Пререквизиты.** Для использования онтологии в образовательных целях, необходимо установить логические связи между понятиями, которые определяют порядок изучения концептов в реальном образовательном про-

цессе по соответствующей национальной учебной программе. Концепт А называется пререквизитом для концепта В, если изучение концепта А, является обязательным для понимания концепта В. Например, понимание концепта *Сложение* требуется для понимания концепта *Умножение*.

Примером цепочки пререквизитов является последовательность концептов онтологии: *Точка* [7-й класс] → *Луч* [7-й класс] → *Угол* [7-й класс] → *Прямой угол* [7-й класс] → *Прямоугольный треугольник* [7-й класс] → *Теорема Пифагора* [8-й класс].

OntoMathEdu предлагает два подхода к определению пререквизитов: прямым путем установления связи между двумя понятиями и, косвенно, путем организации понятий по образовательным уровням.

**Лингвистический уровень онтологии** OntoMath<sup>Edu</sup> состоит из многоязычных словарей (лексиконов), определяющих способы выражения концептов из предметной онтологии в русском, английском и татарском языках. Каждое понятие онтологии имеет имя (слово или словосочетание), значение которого соответствует этому понятию. Каждое понятие снабжается также набором текстовых входов – языковых выражений, отражающих значение понятия, которые извлекаются из математических текстов специальными алгоритмами.

Для представления многоязычных лексиконов онтологии разработана мета-онтологическая модель, которая базируется на существующих онтологиях OntoLex/Lemon (<https://www.w3.org/community/ontolex/>; <https://www.w3.org/2016/05/ontolex/>), LexInfo (<https://lexinfo.net/>) и PreMOn (Predicate Model for Ontologies, <https://premon.fbk.eu/>). Благодаря этому разработанные лексиконы могут быть интегрированы в облако Лингвистических открытых связанных данных (Linguistic Linked Open Data, LLOD, <http://linguistic-lod.org/lod-cloud>). Разработанная мета-онтологическая модель содержит также надстройку над указанными онтологиями, с помощью которой синтаксические фреймы из лексиконов могут быть связаны с материализованными отношениями.

Проектирование онтологии OntoMath<sup>Edu</sup> выполнено в редакторе Protégé.

## Заключение

В статье представлен процесс разработки новой образовательной онтологии OntoMath<sup>Edu</sup> и кратко описана ее текущая версия. Выделены основные проблемы онтологического инжиниринга в образовательной сфере, связанные со спецификой предметного обучения в различных странах в различных образовательных системах, в том числе требующие учета методики преподавания и языка обучения. Онтология OntoMath<sup>Edu</sup> является первой образовательной полилингвальной математической онтологией общего назначения, которая представлена в облаке Открытых связанных данных (Linked Open Data, LOD).

Онтология OntoMath<sup>Edu</sup> находится в процессе дальнейшей разработки, и в ближайших планах применить принципы моделирования, протестированные в настоящем проекте, для разработки других разделов школьной математики.

### Список литературы

- [Barana et al, 2018] Barana A., Di Caro L., Fioravera M., Marchisio M., Rabellino S. Ontology Development for Competence Assessment in Virtual Communities of Practice. / In: Penstein Rosé C., et al. (eds.) // Proceedings of the 19th International Conference Artificial Intelligence in Education (AIED 2018). Part II. LNCS. – Vol. 10948. – P. 9498. – Springer, Cham, 2018. – [https://doi.org/10.1007/978-3-319-93846-2\\_18](https://doi.org/10.1007/978-3-319-93846-2_18).
- [Borgo et al, 2010] Borgo S., Masolo C. Ontological Foundations of DOLCE / In: Poli R., Healy M., and Kameas, A. (eds.) // Theory and Applications of Ontology: Computer Applications. – Springer, Dordrecht, 2010. – P. 279-295. – [https://doi.org/10.1007/978-90-481-8847-5\\_13](https://doi.org/10.1007/978-90-481-8847-5_13).
- [D'Aquin, 2016] D'Aquin M. On the Use of Linked Open Data in Education: Current and Future Practices / In: Mouromtsev D. and D'Aquin M. (eds.) // Open Data for Education: Linked, Shared, and Reusable Data for Teaching and Learning. LNCS. – – Vol. 9500. – Springer, Cham, 2016. – P. 3-15. – [https://doi.org/10.1007/978-3-319-30493-9\\_1](https://doi.org/10.1007/978-3-319-30493-9_1).
- [Dehaye et al, 2016] Dehaye P.-O., Iancu M., Kohlhase M., Kononov A., Lelièvre S., Müller D., Pfeiffer M., Rabe F., Thiéry N.M., Wiesing T. Interoperability in the OpenDreamKit Project: The Math-in-the-Middle Approach / In: Kohlhase M., et al. (eds.) // Proceeding of the 9th International Conference on Intelligent Computer Mathematics (CICM 2016). LNCS. – Vol. 9791. – Springer, Cham, 2016. – P. 117-131. – [https://doi.org/10.1007/978-3-319-42547-4\\_9](https://doi.org/10.1007/978-3-319-42547-4_9).
- [Elizarov et al, 2016] Elizarov A.M., Zhizhchenko A.B., Zhil'tsov N.G., Kirillovich A.V., and Lipachev E.K. Mathematical Knowledge Ontologies and Recommender Systems for Collections of Documents in Physics and Mathematics // Doklady Mathematics. – 2016. – Vol. 93 (2). – P. 231-233. – <https://doi.org/10.1134/S1064562416020174>.
- [Guarino et al, 2009] Guarino N., Oberle D., Staab S. What is an ontology? / In: Staab S., Studer R. (eds.) // International Handbooks on Information Systems. Handbook on Ontologies. – 2th ed. – N.Y. Springer, 2009. – P. 1-17.
- [Guizzardi, 2005] Guizzardi G. Ontological Foundations for Structural Conceptual Model // CTIT PhD Thesis Series. No. 05–74. – Enschede, 2005. – 441 p.
- [Kirillovich et al, 2020a] Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., and Shakirova L. OntoMathEdu: Towards an Educational Mathematical Ontology / In: Edwin Brady, et al. (eds.) // Workshop Papers at 12th Conference on Intelligent Computer Mathematics (CICM-WS 2019). CEUR Workshop Proceedings. – CEUR-WS.org, 2020. – Vol. 2634.
- [Kirillovich et al, 2020b] Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., and Shakirova L. OntoMathEdu: a Linguistically Grounded Educational Mathematical Ontology / In Benzmüller C. and Miller B. (eds.) // Proceedings of the 13th Conference on Intelligent Computer Mathematics (CICM 2020). Lecture Notes in Artificial Intelligence. – Springer, 2020. – Vol. 12236. – P. 157-172.

- [**Kirillovich et al, 2020c**] Kirillovich A., Falileeva M., Nevzorova O., Lipachev E., Dyupina A., Shakirova L. Prerequisite Relationships of the OntoMathEdu Educational Mathematical Ontology / In: J.C. Figueroa-Garc a, et al. (eds.) // Proceedings of the 7th Workshop on Engineering Applications (WEA 2020). Communications in Computer and Information Science. – Springer, 2021. – P. 517-524. – [https://doi.org/10.1007/978-3-030-86702-7\\_44](https://doi.org/10.1007/978-3-030-86702-7_44).
- [**Müller et al, 2017**] Müller D., Gauthier T., Kaliszyc C., Kohlhase M., Rabe F. Classification of Alignments Between Concepts of Formal Mathematical Systems / In: Geuvers H., et al. (eds.) // Proceedings of the 10th International Conference on Intelligent Computer Mathematics (CICM 2017). Lecture Notes in Computer Science. – Springer, Cham, 2017. – Vol. 10383. – P. 83-98. – [https://doi.org/10.1007/978-3-319-62075-6\\_7](https://doi.org/10.1007/978-3-319-62075-6_7).
- [**Nevzorova et al, 2014**] Nevzorova O., Zhiltsov N., Kirillovich A., Lipachev E. OntoMathPRO Ontology: A Linked Data Hub for Mathematics / In: Klinov P. and Mouromstev D. (eds.) // Proceedings of the 5th International Conference on Knowledge Engineering and Semantic Web (KESW 2014). Communications in Computer and Information Science. – Springer, Cham, 2014. – Vol. 468. – P. 105-119. – [https://doi.org/10.1007/978-3-319-11716-4\\_9](https://doi.org/10.1007/978-3-319-11716-4_9).
- [**Paquette et al, 2021**] Paquette, G., Marino, O., Bejaoui, R.: A new competency ontology for learning environments personalization // Smart Learning Environments. – 2021. – Vol. 8 (16). – P. 1-23. – <https://doi.org/10.1186/s40561-021-00160-z>.
- [**Pereira et al, 2018**] Pereira C.K., Matsui Siqueira S.W., Nunes B.P., Dietze S. Linked Data in Education: A Survey and a Synthesis of Actual Research and Future Challenges // IEEE Transactions on Learning Technologies. – 2018. – Vol. 11 (3). – P. 400-412. – <https://doi.org/10.1109/TLT.2017.2787659>.
- [**Tzoumpa et al, 2020**] Tzoumpa D., Karvounidis T., Douligeris C. Circle's Ontology Extended: Circumference and Surface Area of a Circle // Advances in Intelligent Systems and Computing. – 2020. – Vol. 916. – P. 120-132
- [**Муромцев, 2020**] Муромцев Д.И. Модели и методы индивидуализации электронного обучения в контексте онтологического подхода // Онтология проектирования. – 2020. – Т. 10, № 1 (35). – С. 34-49. – <https://doi.org/10.18287/2223-9537-2020-10-1-34-49>.

## ОЦЕНКА КАЧЕСТВА ОБСЛУЖИВАНИЯ КЛИЕНТОВ НА ОСНОВЕ ИХ ЭМОЦИОНАЛЬНОЙ УДОВЛЕТВОРЕННОСТИ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ\*

А.И. Нуриахметов (*nu7530@mail.ru*)

Д.Р. Богданова (*dianochka7bog@mail.ru*)

Н.И. Юсупова (*yusupova.ni@ugatu.su*)

Уфимский университет науки и технологий (УГАТУ), Уфа

В статье предлагается подход к оценке удовлетворенности клиентов на основе распознавания эмоциональных состояний клиентов до и после оказания услуг. Приводится разработанный метод оценки качества обслуживания клиентов, на основе распознавания эмоций. В качестве инструментария применяется метод искусственных иммунных систем для распознавания эмоций по мимике на изображениях. В работе, используя один из методов искусственных иммунных систем, удалось достигнуть максимальной точности в 80 % для задачи распознавания базовых эмоций Пола Экмана. Эти показатели были достигнуты на наборе данных Cohn-Kanade+. Для построения подобной системы были рассмотрены наиболее популярные подходы к распознаванию эмоций на изображениях, а также ключевые концепции типизации эмоций. В предлагаемой модели использовался подход на основе компьютерного зрения, с использованием лицевой разметки. Полученные координаты точек лица были преобразованы в вещественные признаки, а затем, их число было сокращено при помощи метода главных компонент. Разработана автоматизированная система оценки качества обслуживания клиентов на основе их эмоциональной удовлетворенности, позволяющая учитывать эмоциональные состояния клиентов после процесса оказания услуг.

**Ключевые слова:** искусственные иммунные системы, автоматизированный учет эмоций, оценка качества услуг, удовлетворенность клиентов, аффективные вычисления, распознавание эмоций.

---

\* Работа выполнена при финансовой поддержке гранта РФФИ 22-19-00471 и государственного задания № FEUE-2020-0007.

## Введение

Все услуги объединяет один общий фактор, который, является ключевым при выборе поставщика услуг – качество обслуживания клиентов. Главный показатель качества проведенной услуги – удовлетворённость клиента, которая выражается его внутренними чувствами, по отношению к данной услуге. Действительно, нельзя отрицать тот факт, что качество оказания услуги напрямую влияет на эмоциональную удовлетворенность клиента. Например, он может быть раздосадован тем как ему оказали услугу, или же наоборот, очень обрадован её качеством. Тем не менее, чувства клиента являются скрытыми эмоциональными процессами. Узнать о том какие именно чувства испытывает человек можно либо спросив его об этом, либо же изучив его эмоции, распознав их.

До сих пор процесс оценки качества обслуживания клиентов является ручным и не обладает какой-либо системой автоматизации. Для оценки качества компаниям приходится изучать отзывы клиентов, а также просить их проходить различные опросы. Данный процесс можно было бы автоматизировать, используя систему распознавания эмоций, на основе которой можно спроектировать систему оценки качества.

Следовательно, актуальной является задача разработки системы оценки качества обслуживания клиентов на основе их эмоциональной удовлетворенности с использованием современных технологий искусственного интеллекта.

### 1. Обзор существующих подходов к оценке эмоциональной удовлетворенности

Основой исследования эмоциональных состояний человека является теория типизации эмоций. Одной из теорий типизации эмоций является концепция Пола Экмана о шести базовых эмоциях [Ekman et al., 1969; Ekman et al., 1999]. В список базовых эмоций Экмана вошли: страх, удивление, гнев, печаль, отвращение и радость. Весьма важную роль в популяризации концепции типизации эмоций Экмана вносит его “система кодирования действий лица” (от англ. Facial Action Coding System, FACS) [Ekman et al., 1978]. FACS по сути является набором кодировок определенных лицевых мышц человека. Каждой такой комбинации кодов может сопоставляться конкретная эмоция. Данная система нашла себе широкое применение в исследованиях, связанных с задачей распознавания эмоций, и поэтому концепция Экмана является одной из доминирующих концепций типизации эмоций.

За последние десять лет произошел небывалый рост популярности алгоритмов глубокого обучения, их обсуждения даже сумели выйти за рамки научного сообщества, встречаясь все чаще в повседневной жизни. Пожа-

луй, одной из главных тому причин послужила их поразительная способность отлично моделировать сложные, нелинейные функциональные зависимости, что позволило им достигать ранее не достижимых оценок эффективности во множестве задач машинного обучения. Естественно, данный факт также затронул область распознавания эмоций.

При использовании нейронных сетей в задаче распознавания эмоций можно выделить два подхода. В первом нейронные сети используются лишь только для извлечения признаков, необходимых для классификации. Во втором подходе они так же выступают в роли классификаторов. Чаще всего, в обоих подходах можно встретить различные вариации сверточных нейронных сетей. Сверточные нейронные сети зарекомендовали себя лучше всего при работе с распознаванием объектов на изображениях [ImageNet, 2022; LeCun et al., 1998; Badrulhisham et al., 2021; Mehendale et al., 2020; Aslan, 2021]. Тем не менее, они по-прежнему остаются нейронными сетями и сохраняют все присущие им минусы и проблемы, такие как: переобучение, множество параметров, неопределённость необходимых мощностей, большое число обучающих данных. Например, множество варьируемых параметров и неопределенность необходимых мощностей. Также, сверточным сетям требуется большое число обучающих данных, без которых они не смогут выйти на плато своей оптимальной эффективности. А при использовании большого набора обучающих данных, время необходимое на обучение значительно возрастает, требуя высоких вычислительных мощностей. При этом нередкой становится проблема переобучения сети.

Еще одним подходом к распознаванию эмоций можно назвать новый подход с применением искусственных иммунных систем. Действительно, иммунологию можно считать богатым источником для вдохновения при её рассмотрении с вычислительной точки зрения. Например, биологической иммунной системе приходится постоянно решать такие задачи как: распознавание своих и чужих клеток, извлечение признаков антигена, адаптация под новые угрозы, запоминание предыдущих угроз, регуляция своего состояния, а также множество других задач. Между каждой такой биологической задачей можно провести аналогию с задачами машинного обучения [Jerne et al., 1974; Perelson, 1989; Bersini et al., 1990; Ishida et al., 1990; Forrest et al., 1994].

Сегодня можно выделить две большие группы алгоритмов искусственных иммунных систем. А именно – первое поколение алгоритмов, вдохновленное простыми механизмами иммунологии, а также второе поколение, которое использует уже результаты междисциплинарных исследований в данной области. К одному из самых первых видов алгоритмов искусственных иммунных систем можно отнести алгоритмы, построен-



ные на концепции отрицательного отбора [Greensmith et al., 2010]. Основной задачей отрицательного отбора в биологической иммунной системе является обеспечение её лояльности к собственным клеткам. Тем самым иммунная система может находить антигены патогенов и зараженных клеток, при этом не атакуя свои собственные здоровые клетки. Еще одним весьма популярным механизмом иммунологии является теория клональной селекции. Теория клональной селекции служит фундаментом для множества алгоритмов искусственных иммунных систем. В их основе лежат циклические действия оценки аффинности, клонирования, мутации, а также замены антител.

К одному из самых первых алгоритмов данной теории можно отнести алгоритм CLONALG разработанный в 2000 году [De Castro et al., 2000]. Также, существуют и другие весьма популярные алгоритмы, среди которых: AI-NET, B-Cells algorithm (BCA), а также AIRS (Artificial immune recognition system) [Watkins et al., 2004]. В дальнейшем будет применяться модификация алгоритма AIRS.

## **2. Предлагаемый подход к оценке качества оказания услуг**

Качество оказания какой-либо услуги напрямую влияет на внутренние чувства человека. Эмоции могут отражать личностное восприятия качества оказанной услуги. При этом, в отличие от словесного отзыва, эмоции находятся на поверхности и могут быть наблюдаемыми и анализируемы без необходимости в каких-либо дополнительных действиях со стороны клиента. Ему не нужно оставлять письменную запись в книге отзывов, проходить какие-либо опросы, высказывать свое мнению. Эмоции могут быть считаны без доставления какого-либо дискомфорта и напряжения для клиента. Например, можно считать эмоции клиента в момент регистрации на услугу. В таком случае, появляется возможность считать их в момент до оказания услуги и в момент после. А далее, можно произвести сравнение этих эмоциональных состояний и сделать вывод об эмоциональной удовлетворенности клиента качеством оказанной услуги.

Модель предлагаемого процесса оценки качества представлена на рис. 1.

Эмоциональное состояние клиента считывается при регистрации (запись/консультация), а также в момент после получения услуги. На основе сравнения состояний до и после, выставляется соответствующая оценка качества. Таким образом отпадает какая-либо необходимость в дополнительных действиях со стороны клиента. Нет необходимости в ручной обработке данных, система автоматизировано оценивает качество оказанных услуг.

*Рис. 1. Предлагаемая модель процесса оценки качества*

### **3. Математическая постановка задачи оценки качества услуг**

Эмоции одного конкретного человека могут зависеть не только от качества оказываемых услуг, но также и от множества других внешних факторов. Поэтому говорить об оценке качества какой-либо услуги исходя из эмоциональной оценки одного клиента будет некорректно. Для адекватности оценки, имеет смысл рассмотреть её на определенном временном промежутке  $T$ , за который данную услугу пройдет  $m$  клиентов. В таком случае, данную задачу можно кратко представить в виде следующей функции:

$$\{OK_{ij}\}_{i=1}^m \rightarrow OK_j. \quad (3.1)$$

Здесь параметр  $OK_{ij}$  представляет собой оценку качества оказанной  $j$ -ой услуги  $i$ -ому клиенту. Параметр  $OK_j$  представляет из себя финальную оценку для  $j$ -ой услуги на основе данных об  $m$  клиентах, пришедших за промежутки времени  $T$ . Для поиска данной функции, необходимо провести её декомпозицию. Так, рассмотрим задачу поиска оценки качества  $j$ -ой услуги для  $i$ -го клиента. Её можно записать в следующей форме:

$$(K_i, P_j) \rightarrow OK_{ij}. \quad (3.2)$$

Здесь, параметр  $K_i$  это потенциальный клиент организации, который намерен получить услугу. Стоит учесть, что в организациях, зачастую, предоставляют множество разнообразных услуг. Функция оценки качест-

ва не может иметь зависимость только от характеристик клиента. Следовательно, в функции оценки качества появляется параметр  $\Pi_j$ , который представляет собой услугу, которую намерен пройти  $i$ -ый клиент.

Индекс  $i$  представляет из себя порядковый номер клиента, поступившего в систему. В системе не может быть состояния без клиентов, соответственно, данный индекс не может быть равен нулю и начинается с единицы. Рассматривается оценка качества на определенном временном промежутке и число клиентов будет конечным. Индекс  $i$  принадлежит интервалу от единицы до  $m$ .

Аналогично и для индекса  $j$ , который представляет из себя порядковый номер предоставляемой услуги. Очевидно, что число предоставляемых услуг не может быть бесконечным, соответственно, оно является ограниченным сверху. Каждая организация имеет свой определенный набор услуг, ограничим индекс  $j$  некоторым значением  $n$ , которое будет представлять из себя количество услуг, предоставляемых организацией.

Таким образом, получаем пару  $(K_i, \Pi_j)$ , которая характеризует  $i$ -го клиента, получившего  $j$ -ую услугу. Данной паре однозначно соответствует оценка качества  $OK_{ij}$ . Данная пара  $(K_i, \Pi_j)$  является предопределенной информацией, то есть она известная априори.

Проведем декомпозицию данной функции с целью разложить ее на отображения и функции. Входными данными задачи являются изображения с запечатленными на них эмоциями клиента:

$$K_i = \{K_i^{t_1}, K_i^{t_2}, \dots, K_i^{t_l}\}. \quad (3.3)$$

Параметр  $K_i$  в рамках математической модели, представляет из себя массив параметров  $K_i^t$  которые определяют конкретную мимическую эмоцию  $i$ -го клиента в момент времени  $t$ .

Для того чтобы перейти к оценке качества процедуры, изначально необходимо построить классификатор, который позволит выдавать метку класса эмоции для каждого кадра. Рассмотрим следующее отображение:

$$K_i^t \rightarrow \mathcal{E}^i. \quad (3.4)$$

В данном случае  $\mathcal{E}^i$  представляет из себя конкретную метку класса, характеризующую эмоцию  $i$ -го клиента в момент времени  $t$ . Параметр  $K_i^t$  содержит под собой мимическую эмоцию  $i$ -го клиента в момент времени  $t$ . В зависимости от поставленной задачи, а также от используемого метода, данный параметр может раскрываться по-разному. В нашем случае, параметр  $K_i^t$  является изображением, с запечатленным на нем лицом  $i$ -го клиента.

Параметр  $K_i^t$ , как изображение в рамках нашей задачи, содержит излишнюю и не структурированную информацию. Следовательно, к нему необходимо применить специальное преобразование, которое позволило бы сократить число признаков, а также привело их в более удобный для решения формат. В таком случае, обозначим данное преобразование через некую функцию  $E$ , которая принимает на вход мимическую эмоцию  $i$ -го клиента в момент времени  $t$ . Её выходом является конечная совокупность весовых коэффициентов  $x_1^t, x_2^t, \dots, x_n^t$ . Таким образом имеем:

$$E(K_i^t) \rightarrow x_1^t, x_2^t, \dots, x_n^t. \quad (3.5)$$

В итоге, объединив и сократив два предыдущих выражения, получаем задачу классификации, когда известному набору весовых коэффициентов необходимо сопоставить метку определенного класса:

$$x_1^t, x_2^t, \dots, x_n^t \rightarrow \mathcal{E}^t. \quad (3.6)$$

Данное выражение справедливо лишь для некоторого однозначного момента времени  $t$ . Переходя из конкретной временной точки на временной отрезок, для  $i$ -го клиента будем иметь следующее выражение:

$$\begin{array}{ccc} x_1^t, x_2^t, \dots, x_n^t & \mathcal{E}^t & \\ \dots & = & \dots \\ x_1^t, x_2^t, \dots, x_n^t & \mathcal{E}^t & \end{array} \quad (3.7)$$

Таким образом, для  $i$ -го клиента, на некотором временном промежутке  $[0, t_i]$ , получаем вектор состоящий из меток различных классов эмоций, в которых прибывал данный клиент во время мониторинга.

Далее, необходимо провести преобразование данного вектора в оценку эмоции клиента  $\overline{\mathcal{E}}^i$ . Для этого перехода, необходимо построить еще одно отображение, которое будет выполнять необходимые преобразования вектора в конечную оценку. Обозначим данное отображение как некую функцию  $C$ , принимающую в качестве аргумента – вектор меток класса эмоций:

$$C(\mathcal{E}^t, \mathcal{E}^t, \dots, \mathcal{E}^t) \rightarrow \overline{\mathcal{E}}^i. \quad (3.8)$$

Находя оценку эмоций клиента до момента получения услуги и в момент после, можно производить оценку качества данной услуги с точки зрения эмоциональной удовлетворенности клиента  $i$ :

$$(\overline{\mathcal{E}}_0^i, \overline{\mathcal{E}}_n^i) \rightarrow OK_{ij}, \quad i \in N, j \in [1, n]. \quad (3.9)$$

Здесь,  $\overline{\mathcal{E}}_o^i$  – оценка эмоции  $i$ -го клиента до получения услуги,  $\overline{\mathcal{E}}_n^i$  – оценка эмоции  $i$ -го клиента после получения услуги. Оценки эмоций представляют из себя определенный весовой коэффициент, который отражает эмоциональное состояние клиента в данные промежутки времени.

Таким образом, для решения исходной задачи построения системы оценки качества оказания услуг клиентам на основе их эмоциональной удовлетворенности полученными услугами, необходимо найти все представленные преобразования и отображения. Так, функция (3.5) представляет собой этап извлечения признаков, а функция (3.7) является алгоритмом классификации. Отображения (3.8) и (3.9) относятся к процессу выставления оценки системой, на основе определенных эмоциональных состояний клиента.

Переходя к формальной постановке задачи, дано:

- видеоизображения клиентов в моменты до и после получения услуги;
- информация об оказываемых услугах;
- информация о клиенте и о получаемой им услуге;

Необходимо:

- разработать систему распознавания эмоций по мимике на изображениях с использованием искусственных иммунных систем;
- используя разработанную систему распознавания эмоций, спроектировать систему оценки качества обслуживания клиентов на основе их эмоциональной удовлетворенности, а именно: спроектировать серверное приложение – платформу, позволяющую производить мониторинг оценок качества, а также контроль и учет информации о компании (услуги, поставщики, филиалы); спроектировать клиентское приложение, которое будет производить оценку качества обслуживания клиентов, используя систему распознавания эмоций.

#### 4. Выставление оценки качества

На данном этапе:  $\overline{\mathcal{E}}^o$  – оценка эмоций клиента за промежуток до получения услуги;  $\overline{\mathcal{E}}^n$  – оценка эмоций клиента за промежуток после получения услуги.

Каждая такая оценка характеризуется парой  $\{\overline{\mathcal{E}}_n^i, \overline{\mathcal{E}}_p^i\}$ , где  $\overline{\mathcal{E}}_n^i$  – коэффициент негативных эмоций, а  $\overline{\mathcal{E}}_p^i$  – коэффициент позитивных. Коэффициенты вычисляются как отношения к общему числу эмоций.

Обозначим  $D = (\overline{\mathcal{E}}_p^n - \overline{\mathcal{E}}_n^n) - (\overline{\mathcal{E}}_p^o - \overline{\mathcal{E}}_n^o)$ , тогда оценка качества находится по следующей формуле:

$$\begin{array}{ll}
|D| \leq M, & OK = NC \\
OK = D > M, & OK = P \\
D < M, & OK = N
\end{array} \quad (4.1)$$

где  $NC$  – нейтральная оценка (не существенное изменение),  $P$  – позитивная оценка,  $N$  – негативная оценка.  $M$  – параметр определяющий существенность изменения эмоционального состояния.

Предлагаемая схема оценки качества оказания услуг представлена на рис. 2.

*Рис. 2. Схема процесса оценки качества*

Разрабатываемая система автоматически ведет подсчет отрицательных и положительных эмоций клиента как в момент до прохождения услуги, так и после. Система оценивает соотношения отрицательных и положительных эмоций в процентах в момент до и в момент после оказания услуг и выводит получившуюся разницу оценок. Если после оказания услуг разница является “значительно” больше разницы до получения услуг, то ставится положительная оценка, иначе – отрицательная. Если же изменение произошло незначительное, не превышающее 10% как в положительную, так и отрицательную сторону, то в таком случае ставится нейтральная оценка и фиксируется, что не было существенного изменения в эмоциональном состоянии клиента после оказания услуг.

## 5. Предлагаемая модель системы распознавания эмоций

Предлагаемый подход к распознаванию эмоций состоит из трех этапов: на первом этапе происходит обработка изображения, а также извлечение лицевых координат; на втором этапе проходит преобразование координат в признаки для классификации; на третьем этапе проводится классификация эмоций клиента с помощью модифицированного алгоритма искусственной иммунной системы. Общая схема модели изображена на рис. 3.

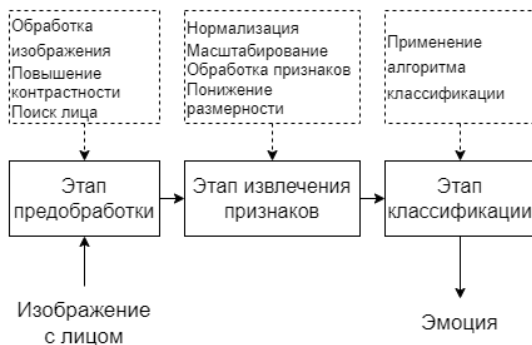


Рис. 3. Общая схема модели

Первым этапом является этап предобработки данных. Изначально на вход модели подаются сырые данные – изображения с запечатлёнными на них лицами. Так, происходит преобразование изображения в черно-белые оттенки, повышение контрастности. На данном этапе реализуются алгоритмы поиска лица на изображении, обрезания области лица, его нормализации.

Следующим этапом модели распознавания эмоций является этап извлечения признаков. Здесь происходит извлечение признаков, подходящих для распознавания из уже обработанного изображения, их нормализация, стандартизация, и применяются алгоритмы понижения размерности признакового пространства.

Последнем этапом в модели является этап классификации. На данном этапе, извлечённые и обработанные признаки поступают в алгоритм классификации, результатом которого уже является фактическая, конкретная метка класса – определенная эмоция.

## 6. Алгоритмы классификации

Существует широкое множество всевозможных алгоритмов для решения задачи классификации объектов. Авторы в [Нуриахметов и др., 2021] приводят анализ применимости алгоритмов искусственных иммунных

систем к распознаванию эмоций человека. В разрабатываемой системе в качестве алгоритма классификации используется алгоритм искусственной иммунной системы распознавания AIRS [Watkins et al., 2004].

Данный алгоритм относится к теории искусственных иммунных систем и принадлежит классу алгоритмов клональной селекции. AIRS является одним из первых алгоритмов теории искусственных иммунных систем, который изначально подразумевался для решения всевозможных задач классификации.

В основе данного алгоритма лежит идея иммунной системы по взаимодействию В-лимфоцитов и антигенов. В качестве основного строительного блока в алгоритме выступают В-лимфоциты. Их можно подразделить на две категории: искусственные иммунные шары распознавания (ARB – artificial recognition ball) и на клетки памяти. ARB – представляют из себя группу одинаковых В-лимфоцитов, созданных с целью сократить возможные повторения на этапе обучения. Изначально идея ARB была предложена в алгоритме “Artificial Immune Network algorithm” (AINE) [Knight et al., 2021]. Клетки памяти же используются уже на этапе классификации. Каждый В-лимфоцит является вещественным вектором с признаками и меткой класса, который представляет из себя потенциальное решение задачи. Каждый объект, подлежащий распознаванию, представляет из себя потенциальный антиген, который по сути является вещественным вектором признаков.

В работе алгоритма вектор-признаков – это вектор вещественных значений, каждое из которых является признаком рассматриваемого объекта. Мера сходства – это измерение близости двух антител или антигенов. В рамках данной модели под близостью будем понимать евклидово расстояние между векторами признаков данных двух объектов. Порог сходства (AT – affinity threshold) – это средняя мера сходства между всеми антителами или антигенами в рассматриваемой выборке. Скаляр порога сходства (ATS – affinity threshold scalar) – это срез порога сходства, то есть порог сходства, умноженный на некоторый коэффициент в диапазоне от 0 до 1. Антитело – это пара из вектора признаков и метки класса, т.е. объект обучения. Антиген – это пара из вектора признаков и метки класса, т.е. элемент из обучающей выборки, подлежащий распознаванию. Стимуляция антитела – это разность единицы и степени сходства антитела и рассматриваемого антигена. Порог стимуляции - это средний порог сходства обучающего набора. Искусственный иммунный шар распознавания (ARB – Artificial Recognition Ball) является подобием реального В-лимфоцита, содержит в себе выборку антител, а также число ресурсов, которыми он обладает, и свой уровень стимуляции. Клетка памяти (MC – Memory Cells) – это антитело с наибольшим коэффициентом стимуляции по отношению к рассматриваемому антигену. Кандидат в клетки памяти (CMC – candidate Memory Cell) – это антитело, метка класса которого совпадает с



меткой класса рассматриваемого антигена, а также уровень стимуляции которого является наибольшим по отношению к данному антигену. Коэффициент клонирования – это коэффициент, позволяющий определить дозволенное число клонов для рассматриваемого шара распознавания. Коэффициент гипермутации – это коэффициент определения числа клонов, которое может быть добавлено клеткой памяти в общую популяцию антител. Коэффициент мутации – это вероятность мутации конкретного признака из вектора признаков. Ресурсы – это параметр, контролирующий максимальное число шаров распознавания в системе.

Изначальная версия алгоритма набрала большую популярность и подверглась тщательному анализу и изучению, результатом которого стала более усовершенствованная версия алгоритма – AIRS2. Основным отличием второй версии алгоритма является то, что в процессе обучения алгоритма, пул ARB клеток очищается после каждого рассмотренного антигена.

В общем случае жизненный цикл алгоритма можно представить следующим образом (рис. 4).

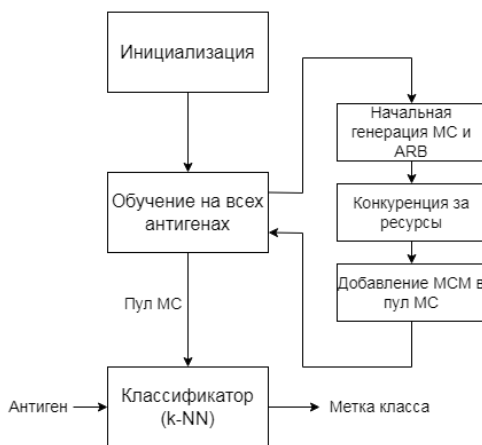


Рис. 4. Схема алгоритма AIRS

Как видно из схемы, AIRS содержит в себе четыре стадии: инициализация, идентификация МС и генерация ARB, конкуренция за ресурсы и разработку кандидата в МС, добавление МС в общий пул.

На этапе инициализации происходит первичная предобработка данных и инициализация стартовых значений. Происходит нормализация признаков обучающего набора, чтобы мера сходства между признаками не выходила за рамки отрезка от 0 до 1. Затем, вычисляются значения порога сходства по формуле

$$AT = \frac{\sum_{i=1}^n \sum_{j=i+1}^n affinity(ag_i, ag_j)}{\frac{n(n-2)}{2}}, \quad (6.1)$$

где  $ag$  – антигены из обучающей выборки,  $affinity$  – функция вычисления сходства.

Формируется начальная популяция коллекций MC и ARB. Данный процесс выполняется путем случайного выбора антигенов из обучающего набора в количестве от 0 и более штук.

Затем, происходит начальная популяция коллекций MC и ARB. Данный процесс выполняется путем случайного выбора антигенов из обучающего набора в количестве от 0 и более штук.

Процесс обучения алгоритма представлен на рис. 5.

На втором этапе, для каждого антигена из обучающего набора вычисляются степени стимуляции каждой клетки памяти имеющий одинаковый класс с антигеном. Клетка имеющая наибольший показатель стимуляции используется для генерации новых клеток ARB с применением мутаций. Число генерируемых клеток мутации пропорциональной степени сходства выбранной клетки памяти и рассматриваемого антигена. Мутация применяется на каждый компонент вектора признаков данной клетки с определенным шансом, который диктуется коэффициентом мутации.

Далее, на третьем этапе происходит генерация лучшего кандидата в клетки памяти для рассматриваемого антигена. Каждому ARB из пула выдается определенное число, называемое ресурсами, определяемое по следующей формуле:

$$r = ClonalRate \cdot Stim(ARB, ag), \quad (6.2)$$

где  $ClonalRate$  – коэффициент мутации,  $Stim$  – функция нахождения стимуляции между антителом ARB и антигеном  $ag$ .

Если число необходимых ресурсов начинается превосходить число доступных, то из популяции ARB удаляются те клетки, которые имеют наименьший уровень стимуляции и их ресурсы освобождаются. Далее происходит проверка на критерий останки. Если порог сходства ( $AT$ ) всех клеток ARB с рассматриваемым антигеном превосходит средний порог сходства обучающего набора, то происходит переход на следующую итерацию обучения, то есть в рассмотрение берется новый антиген. Иначе, ARB клетки будут подвержены мутации, клональной экспансии и повторной битве за ресурсы до тех пор, пока критерий не будет достигнут. Наконец, ARB клетка с наилучшим показателем стимуляции будет выбрана в качестве СМС клетки.

На четвертом этапе принимается решение о добавлении нового кандидата МС клетки в общий пул долгоживущих МС клеток. Данное решение принимается на основе сравнения степени стимуляции кандидата и наилучшей клетки памяти для данного антигена. Если степень стимуляции кандидата превосходит степень стимуляции выбранной клетки памяти, то кандидат заменяет клетку памяти в общем пуле клеток памяти. После проведения процесса обучения, финальный набор МС клеток можно использовать для проведения классификации с использованием метода k-NN и большинства голосов [Brownlee et al., 2005].

## 7. Тестирование и анализ модели

Для обучения и тестирования построенной модели использовался набор данных Cohn-Kanade (СК+) [Lucey et al., 2010]. Данный набор является весьма популярным что позволяет сравнивать полученные результаты с

результатами других исследований. В своей основе он содержит 7 базовых эмоций Пола Экмана. Набор содержит порядка 40% лиц мужского пола, а также порядка 60% лиц женского пола. Все лица находятся в возрастном диапазоне от 18 до 50 лет. Общее количество лиц превышает число в 200 человек. Сам же набор содержит порядка 1000 изображений размера 48 на 48 пикселей. Данного числа может оказаться недостаточно для обучения нейронной сети, но вполне достаточно для обучения предлагаемой модели. Для вычислительного эксперимента в набор данных вошли 3 эмоции, а именно две негативные эмоции – отвращение и гнев, и одну позитивную – счастье. Данное ограничение сделано в силу того, что другие эмоциональные проявления являются менее информативными по отношению к решаемой проблеме. Для тестирования точности будем использовать методику стратифицированной кросс-валидации по K блокам, с параметром равным 10.

Параметры алгоритма Artificial immune recognition system (AIRS2): коэффициент  $k = 5$  (для kNN), коэффициент гиперклонирования HCR = 25, коэффициент мутации MT = 0.1, максимальное число ресурсов MR = 50, коэффициент клонирования CR = 1, коэффициент начальной инициализации клеток памяти MIR = 0.4, порог сходства AT = 0,7.

В результате эксперимента была получена точность алгоритма 80%, скорость обучения 10 секунд, скорость классификации одного объекта: 0,01 секунды.

## **8. Разработка системы автоматизированной оценки качества обслуживания клиентов**

На этапе проектирования системы автоматизированной оценки качества обслуживания клиентов была разработана функциональная модель процесса работы информационной системы, изображенная на рис. 6.

Основными этапами работы информационной системы являются: начальная инициация приложения; распознавание эмоционального состояния клиента; фиксирование в памяти информации о клиенте, его эмоционального состояния, информации об услуге; повторное распознавание эмоционального состояния клиента; оценка качества; отправка информации о качестве услуги и клиенте на сервер; мониторинг оценок качества.

Клиентское приложение участвует в этапах распознавания эмоционального состояния, фиксирования в памяти информации о клиенте, повторного распознавания эмоционального состояния, а также в этапах вычисления оценки качества и её последующей отправки на сервер. Серверное приложение участвует лишь только в последнем этапе, позволяя сотрудникам просматривать информацию о клиентах, сотрудниках, оценках и т.д.

*Рис. 6. Процесс работы с использованием информационной системы*

Структура разработанной информационной системы состоит из двух приложений, образующих программный комплекс и представлена на рис. 7.

*Рис. 7. Структура информационной системы оценки качества услуг*

Первым приложением является серверное приложение – платформа, позволяющая производить мониторинг выставляемых оценок качества, а также вести учет и контроль услуг организации, а также её филиалов и сотрудников, оказывающих данные услуги. В качестве архитектуры платформы была выбрана чистая архитектура, а в качестве основы методология предметно-ориентированного проектирования. Для разработки платформы был выбран набор технологий платформы .NET, а именно: инфраструктура ASP.NET Core, ORM система Entity Framework Core, СУБД MS SQL. Для проектирования клиентской части платформы использовались технологии библиотек JQuery и Bootstrap.

Клиентское приложение было спроектировано таким образом, что всю информацию о предоставляемых услугах, а также сотрудниках оно получает из серверного приложения, путем использования его API интерфейса. Также, оценки качества из клиентского приложения отправляют на серверное приложение, что позволяет производить мониторинг оценок в реальном времени, без необходимости ручного скачивания или переноса.

## **Заключение**

Проведен анализ существующих подходов к оценке эмоциональной удовлетворенности клиентов. Рассмотрена дискретная теория типизации эмоций Пола Экмана. Для проектирования системы распознавания эмоций по мимике человека были проанализированы основные подходы: система кодирования лицевых движений Пола Экмана; использование сетей глубоко обучения; использование искусственных иммунных систем. В качестве признакового пространства был выбран подход с использованием лицевой разметки по 68 ключевым точками. К данному подходу была разработана методика предобработки изображения, включающая этапы коррекции контрастности, корректировки угла наклона головы, обрезания изображения и его масштабирования и сокращения размерности пространства признаков.

Разработана математическая модель задачи оценки качества обслуживания клиентов на основе их эмоциональной удовлетворенности. В предлагаемой системе распознавания эмоций в качестве алгоритма классификации был использован алгоритм клональной селекции. AIRS2 из теории искусственных иммунных систем. Было проведено тестирование предлагаемой модели по различным показателям.

Точность системы распознавания эмоций превышает 80%, а скорость классификации одной эмоций не превышает 0,002 секунды. Для обучения системы распознавания эмоций использовался набор данных СК+, а для оценок эффективности алгоритма использовалась методика k-блочной кросс-валидации, с параметром k равным 10.

Спроектирована и разработана автоматизированная система оценки качества обслуживания клиентов на основе их эмоциональной удовлетворенности с использованием технологий искусственных иммунных систем. Было проведено контрольное тестирование качества, разработанного ПО с целью установить его метрические оценки. По результатам тестирования был сделан вывод о том, что программный продукт удовлетворяет основным требованиям, а также обладает стабильностью работы, позволяющей внедрять разработанный продукт.

### Список литературы

- [Aslan, 2021] Aslan M. CNN based efficient approach for emotion recognition // Journal of King Saud University-Computer and Information Sciences. – 2021.
- [Badrulhisham et al., 2021] Badrulhisham N.A.S., Mangshor N.N.A. Emotion Recognition Using Convolutional Neural Network (CNN) // Journal of Physics: Conference Series. – IOP Publishing, 2021. – Vol. 1962, No. 1. – P. 012040.
- [Bersini et al., 1990] Bersini H., Varela F.J. Hints for adaptive problem solving gleaned from immune networks // International Conference on Parallel Problem Solving from Nature. – Springer, Berlin, Heidelberg, 1990. – P. 343-354.
- [Brownlee et al., 2005] Brownlee J. et al. Artificial immune recognition system (airs)-a review and analysis // Center for Intelligent Systems and Complex Processes (CISCP), Faculty of Information and Communication Technologies (ICT), Swinburne University of Technology, Victoria, Australia, 2005.
- [De Castro et al., 2000] De Castro L.N., Von Zuben F.J. The clonal selection algorithm with engineering applications // Proceedings of GECCO. – 2000. – Vol. 2000. – P. 36-39.
- [Ekman et al., 1969] Ekman P., Sorenson E.R., Friesen W.V. Pan-cultural elements in facial displays of emotion // Science. – 1969. – Vol. 164, No. 3875. – P. 86-88.
- [Ekman et al., 1978] Ekman P., Friesen W. V. Facial action coding system // Environmental Psychology & Nonverbal Behavior. – 1978.
- [Ekman et al., 1999] Ekman P. Basic emotions // Handbook of cognition and emotion. – 1999. – Vol. 98, No. 45-60. – P. 16.
- [Forrest et al., 1994] Forrest S. et al. Self-nonsel self discrimination in a computer // Proceedings of 1994 IEEE computer society symposium on research in security and privacy. – Ieee, 1994. – P. 202-212.
- [Greensmith et al., 2010] Greensmith J., Whitbrook A., Aickelin U. Artificial immune systems // Handbook of Metaheuristics. – Springer, Boston, MA, 2010. – P. 421-448.
- [ImageNet, 2022] ImageNet [Электронный ресурс] // ImageNet. – Режим доступа: <https://www.image-net.com/ImageNet> (дата обращения: 10.02.2022).
- [Ishida et al., 1990] Ishida Y. Fully distributed diagnosis by PDP learning algorithm: towards immune network PDP model // 1990 IJCNN International Joint Conference on Neural Networks. – IEEE, 1990. – P. 777-782.
- [Jerne et al., 1974] Jerne N.K. Towards a network theory of the immune system // Ann. Immunol. – 1974. – Vol. 125. – P. 373-389.

- [**Knight et al., 2021**] Knight T., Timmis J. AINE: An immunological approach to data mining // Proceedings 2001 IEEE International Conference on Data Mining. – IEEE Computer Society, 2001. – P. 297-304.
- [**LeCun et al., 1998**] LeCun Y. et al. Gradient-based learning applied to document recognition // Proceedings of the IEEE. – 1998. – Vol. 86, No. 11. – P. 2278-2324.
- [**Lucey et al., 2010**] Lucey P. et al. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. – IEEE, 2010. – P. 94-101.
- [**Mehendale et al., 2020**] Mehendale N. Facial emotion recognition using convolutional neural networks (FERC) // SN Applied Sciences. – 2020. – Vol. 2, No. 3. – P. 1-8.
- [**Perelson, 1989**] Perelson A.S. Immune network theory // Immunol. Rev. – 1989. – Vol. 110, No. 5. – P. 36.
- [**Watkins et al., 2004**] Watkins A., Timmis J., Boggess L. Artificial immune recognition system (AIRS): An immune-inspired supervised learning algorithm // Genetic Programming and Evolvable Machines. – 2004. – Vol. 5, No. 3. – P. 291-317.
- [**Нуриахметов и др., 2021**] Нуриахметов А.И., Богданова Д.Р. Распознавание эмоций на изображениях и искусственные иммунные системы // Моделирование, оптимизация и информационные технологии. – 2021. – Т. 9, №. 3. – С. 7-8.



## ОТ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ К ПОЗНАНИЮ ПРИРОДЫ\*

В.Г. Редько (*vgredko@gmail.com*)  
НИИ системных исследований РАН, Москва

В работе анализируется подход к исследованию процессов происхождения способов научного познания на основе моделей когнитивных агентов. Подчёркивается важность внутренних моделей автономных агентов, а также процессов предсказания агентами. Такие внутренние модели агентов и процессы предсказания используются в обучении с подкреплением. Эти внутренние модели могут рассматриваться как предшественники моделей природы, используемых в научном познании.

**Ключевые слова:** обучение с подкреплением, когнитивный агент, познание природы.

### Введение

В настоящей работе анализируется следующий вопрос: как, отталкиваясь от методов обучения с подкреплением, идти к построению и изучению моделей когнитивных автономных агентов, познающих закономерности природы. Выделяются методы обучения с подкреплением, которые могут быть использованы при изучении когнитивных автономных агентов, познающих закономерности природы.

### 1. Методы обучения с подкреплением и когнитивные автономные агенты

В этом разделе выделим ряд когнитивных способностей автономных агентов, используемых в обучении с подкреплением [Sutton et al., 2018]. В обучении с подкреплением часто используются внутренние модели агентов, в которых запоминаются полученные ранее данные о взаимодействии агента с внешней средой: запоминаются ситуации, действия

---

\* Настоящая работа выполнена в рамках государственного задания по проведению фундаментальных научных исследований по теме "Исследование нейроморфных систем обработки больших данных и технологии их изготовления" № FNEF-2022-0003.

агента в этих ситуациях и получаемые при этом подкрепления. В результате чего обучение агента получать положительные подкрепления происходит значительно быстрее, чем при использовании этих данных непосредственно в процессе поведения агентов без использования запоминания предыдущего опыта. Хотя просто запоминание всех данных не всегда эффективно. Поэтому часто используется иерархический подход при построении внутренних моделей агентов в обучении с подкреплением [Панов, 2022; Dzhivelikian et al., 2022].

Отметим, что в работе [Dzhivelikian et al., 2022] построена схема внутренней мотивации поведения. То есть, даже при отсутствии обычной мотивации (связанной, например, с потребностями агента или поставленными целями) к активному поведению у агента возникает сигнал внутренней мотивации, который стимулирует активность агента в отсутствие внешнего сигнала.

Отметим, что потребность агента (или биологического организма) к накоплению знаний подчёркивается в работах А.А. Жданова [Жданов, 2008], т.е. потребность накопления знаний о внешнем мире рассматривается, как одна из ведущих потребностей организма.

В работе [Schmidhuber, 2015] разработан подход к построению и использованию внутренней системы управления агента, состоящей из двух нейронных сетей: Модели и Контроллера. Система управления используется в обучении с подкреплением. Модель запоминает полезную информацию, которая служит для получения максимальных подкреплений, причём Модель эффективно сжимает информацию о предыдущих событиях. Также в этой работе отмечается, что с помощью Модели можно исследовать внешний мир за счёт внутренней награды. Отмечается, что агента, использующего Модель, с помощью которой происходит изучение внешнего мира, можно рассматривать как предшественника учёного, познающего природу (artificial scientist). Последняя идея близка к подходу работы [Red'ko, 2012], в которой предлагаются принципы функционирования агента-физика, познающего законы природы аналогично известным учёным (немного подробнее об этом см. следующий раздел).

В работе [Schmidhuber, 2018] предлагается использовать в рекуррентной нейронной сети (используемой при обучении с подкреплением) запоминание и использование предыдущего опыта, т.е. запоминание нового опыта без потерь информации о старом опыте. Отмечается, что это возможно аналогично накоплению знаний в мозге человека; действительно, ведь у нас постепенно запоминается всё новая и новая информация без существенного забывания старой. Также в этой работе отмечается, что при этом возможно автоматическое упорядочение решаемых проблем.

Ещё одно интересное направление, близкое к обучению с подкреплением – модели обобщающих агентов, при этом один обобщающий агент может выполнять целый спектр различных задач [Reed et al., 2022].

Таким образом, обучение с подкреплением использует модели когнитивных автономных агентов, обладающих нетривиальными познавательными способностями.

Отметим, что интересный иерархический подход в эволюционном контексте, а именно, концептуальная теория метасистемных переходов, был предложен В.Ф. Турчиным, который в книге «Феномен науки» рассматривает биологическую эволюцию с кибернетической точки зрения, а эволюцию научного познания – как продолжение биокибернетической эволюции [Turchin, 1977; Турчин, 1993]. Кратко суть теории метасистемных переходов сводится к следующему: переход от нижних уровней системной иерархии к верхним происходит путем метасистемных переходов. Каждый метасистемный переход можно рассматривать как объединение ряда подсистем  $S_i$  нижнего уровня и появление дополнительного механизма управления  $C$  объединенными подсистемами. В результате метасистемного перехода формируется система  $S'$  нового уровня ( $S' = C + \sum_i S_i$ ), которая может быть включена как подсистема в следующий метасистемный переход.

Турчин рассматривает метасистемный переход как некий кибернетический аналог физического фазового перехода. Он уделяет особое внимание количественному накоплению «потенциала развития» в подсистемах  $S_i$  перед метасистемным переходом на качественно новый уровень иерархии, а также процессу размножения и развития подсистем предпоследнего уровня иерархии после метасистемного перехода.

## **2. Свойства, которые целесообразно заложить в модели агентов, познающих природу**

В работе [Red'ko, 2012] были предложены принципы построения модели агента-физика, который мог бы аналогично И. Ньютону самостоятельно изучать природу. Кратко перечислим способности такого агента, самостоятельно познающего природу. У агента должна быть база знаний, должно быть стремление к получению новых знаний и к обобщению знаний, должна быть любознательность, направляющая агента к постановке вопросов о внешнем мире и решению этих вопросов путем постановки экспериментов. Агент должен учитывать многочисленные связи между уже имеющимися знаниями. Должен быть коллектив агентов, исследующий внешний мир, и должны быть коммуникации между агентами (если основной агент – аналог И. Ньютона, другие агенты могут быть аналогами Г. Галилея, И. Кеплера, Р. Декарта, Р. Гука, Г.В. Лейбница). Должно быть

самосознание агентов, эмоциональная самооценка результатов своей деятельности и стремление агента достигнуть наиболее высоких результатов в коллективе агентов. Агенты должны иметь стремление к получению наиболее ясных, чётких и компактных знаний, таких как законы Ньютона или аксиомы Евклида. И, конечно, агенты должны освоить возможности логических выводов, позволяющих получить многочисленные следствия законов и аксиом. Понятно, что путь к построению модели познающего агента такого уровня весьма непрост, но, все же, он просматривается.

Рассмотрим и другие свойства, которые целесообразно заложить в познающего агента.

Как уже было сказано, важное понятие – внутренняя модель агента, с помощью которой он формирует представления о внешнем мире и использует знания, сложившиеся в этой модели при взаимодействии с внешним миром. Такие модели аналогичные научным теориям. Более того, общая научная картина мира – создаваемая всем международным научным сообществом – также может рассматриваться как совокупность моделей. Используя научные модели, мы можем делать предсказания будущих событий во внешнем мире.

В работе [Редько и др., 2021] анализировались подходы к формализации процессов озарения у животных и человека. При этом подчёркивалось, что при озарении большую роль играет подсознательная память, с помощью которой внезапно находится решение достаточно долго обдумываемой проблемы, после чего решение переносится в рабочую память, в сознание. В определенном смысле рассмотренные подсознательные и сознательные процессы сходны с процессами в Системе 1 и в Системе 2 в понимании Канемана [Канеман, 2013].

В работе [Laukien, 2016] была предложена интересная иерархическая адаптивная архитектура, машина Феймана, которая представляет собой иерархию простых нейродинамических модулей, обучающихся адаптивно предсказывать свои состояния. Каждый модуль машины Феймана является одновременно кодером входного сигнала из предыдущего уровня иерархии и декодером предсказания более высокого уровня, которые совместно предсказывают значения активаций на следующем шаге времени.

С.А. Шумский предложил обобщение машины Феймана, а именно схему иерархического управления адаптивным поведением агента, в которой на каждом уровне пара кодер-декодер дополняется примитивным планирующим механизмом, способным комбинировать найденные кодером символичные представления, запоминать типовые комбинации символов и выбирать из них наиболее перспективные паттерны поведения. Подчеркнём, что в этой схеме одновременно рассматриваются уровни планирования в разных временных масштабах, что позволяет верхним

этапам планирования использовать отдаленные цели, подчинив текущее поведение долгосрочным интересам агента. Подробнее см. главу 6 в книге [Шумский, 2019]. На основе этой схемы была предложена модель искусственной психики ADAM (Adaptive Deep Autonomous Machine), реализующая иерархическую архитектуру глубокого обучения с подкреплением [Шумский, 2022].

### 3. Начальные модели агентов, познающих внешний мир

В работе [Редько, 2020] построена и исследована модель эволюции популяции самообучающихся агентов, в которой агенты запоминают причинно-следственные связи между ситуациями внешней среде. Познание причинно-следственных связей – одно из важных познавательных свойств человека. Но как это свойство могло бы возникнуть? В указанной работе сделана попытка подойти к исследованию этого вопроса. Модель построена, отталкиваясь от рассуждений Дэвида Юма. В «Исследовании о человеческом познании» (1748 г.) Юм подверг сомнению использование человеком понятия причинной связи [Юм, 1748/1966]. А именно, он задался вопросом: почему, когда мы видим, что за одним явлением *A* постоянно следует другое *B*, то мы приходим к выводу, что *A* является причиной *B*? Например, когда мы многократно наблюдаем, что Солнце освещает камень, и камень нагревается, то мы говорим, что солнечный свет есть причина нагревания камня. И почему мы можем делать вывод тогда, когда мы наблюдаем такую пару последовательных событий много раз, но не делаем вывод о причинно-следственной связи, когда мы только один раз наблюдаем такую пару? Фактически Юм задался вопросом: что заставляет нас делать выводы о происходящих в природе явлениях? Что лежит в основе этих выводов? Юм попытался понять, откуда мы берем основание заключать, что *A* есть причина *B*. Он посмотрел на этот вопрос, как он пишет, со всех сторон и не нашел никакого другого основания, кроме некоторого внутреннего чувства привычки. То есть, имеется какое-то наше внутреннее свойство, которое заставляет нас утверждать, что если за *A* постоянно следует *B*, то *A* есть причина *B*. И это внутреннее чувство заставляет нас после того, когда мы сделали такое умозаключение, и снова видим событие *A*, ожидать, что за *A* вновь последует и событие *B*.

В работе [Редько, 2020] проанализирована компьютерная модель внутреннего чувства привычки у автономных агентов. Рассматривалась популяция агентов, попадающих в разные ситуации. Предполагалось, что в основном ситуации случайные и чередуются во времени случайно. Но возможны и причинные связи между ситуациями. Считалось, что эпизодически появляются пары последовательно чередующихся ситуаций:

$$S_{1i}(t) \rightarrow S_{2j}(t+1),$$

где  $t$  – дискретное время,  $t = 1, 2, \dots$  Эти пары представляют собой причинно-следственные связи мира. Некоторые из ситуаций-следствий  $S_{2j}$  (вторые элементы пар) существенны для агентов. В этих ситуациях агенты могут приобретать или терять существенную часть своего ресурса  $R(t)$ . Эти ситуации  $S_{2j}$  были благоприятными, если агент приобретает в них ресурс, и неблагоприятными, если в такой ситуации агент теряет ресурс.

Считалось, что имеется два вида агентов: агенты первого вида и агенты второго вида. Агенты первого вида не имеют чувства причинности, агенты второго вида имеют чувство причинности.

Агент с чувством причинности при обучении запоминает в своей базе знаний причинно-следственные связи между повторяющимися многократно парами ситуаций  $S_{1i} \rightarrow S_{2j}$  и после обучения заранее готовится к появлению благоприятных и неблагоприятных ситуаций  $S_{2j}$ . Путём компьютерного моделирования было показано, что агенты с чувством причинности явно используют это чувство. Их ресурс растёт быстрее, чем у агентов без чувства причинности. Агенты с большим ресурсом чаще рожают потомков, поэтому в результате эволюции агенты с чувством причинности обычно вытесняют из популяции агентов без чувства причинности.

Таким образом, в работе [Редько, 2020] путём компьютерного моделирования продемонстрировано, что агенты с чувством причинности преимущественно вытесняют других агентов из популяции.

В ряде разработанных нами моделей агенты имели мотивации к выполнению тех или иных действий (см. [Редько, 2018], глава 5).

На примере моделирования многозвенного целенаправленного поведения новокаледонских ворон исследовалась модель формирования воронной-агентом планов цепочки действий, составленных из отдельных звеньев [Red'ko et al., 2016].

В моделях поискового поведения рыб в лабиринтах вводилась уверенность агентов в получаемых знаниях, в получаемых предсказаниях будущих ситуаций. Полагалось, что модельные агенты должны находить баланс между уверенностью в своих знаниях и поиском новых знаний, если уверенность в достигнутых знаниях высока [Red'ko et al., 2015].

Отметим, что поисковое поведение агентов, как модельных организмов, связано с нетривиальным переключением между мотивациями. Например, пусть у агента имеется пищевая потребность и потребность поиска, изучения мира. В процессе функционирования агента происходит переключение между ведущими мотивациями, т.е. переключение между видами действий агента. Как подчеркивается в работах биологов [Непом-

нящих., 2012; Непомнящих., 2013; Непомнящих., 2014], в поведении животных постоянно присутствуют две противоположные тенденции, не связанные непосредственно с физиологическими потребностями. Одна из них – поиск новой, непредсказуемой стимуляции, а другая – стремление четко предсказывать результаты своего поведения. В биологических исследованиях было замечено, что если животное сыто (пищевая мотивация не слишком высока), то в его поведении присутствует стремление найти новые объекты или обнаружить новые свойства у известных предметов. Для объяснения такого поведения была выдвинута гипотеза «уменьшения неопределенности» [Inglis et al., 1997; Inglis et al., 2001], которая предполагает наличие у животного постоянной мотивации к сбору информации о среде. Эта исследовательская мотивация преобладает, если основные потребности (например, в пище) удовлетворены. Тогда поведение животного, стимулируемого мотивацией к сбору информации об окружающей среде, приводит к уменьшению неопределенности. Если же животное сильно голодно и видит пищу, то исследовательское поведение отсутствует.

Важная возможность пополнения знаний – использование латентного обучения, т.е. обучения, при котором животное или агент не получает сразу награды, тем не менее, исследует новые для него ситуации, исследует природу [Зорина и др., 2001]. Полученные при этом знания могут использоваться позднее, эти знания облегчают решение требуемых задач. Пример моделирования эффекта латентного обучения при поиске награды в двумерном лабиринте приведен в работе [Witkowski, 2007].

Обобщая свойства агентов в отмеченных биологически инспирированных моделях, можно выделить следующие характерные черты познавательных свойств и внутренних моделей исследованных когнитивных агентов (см. [Редько, 2018], глава 7):

1) Оценка *знаний* о каждой из возможных ситуаций.

2) Движение к ситуации, знания о которой минимальны. Тем самым *агент может самостоятельно формировать новые текущие цели своего поведения*. Например, рыба в лабиринте может решить двигаться в тот коридор, для которого знания минимальны, т.е. в тот коридор, который она давно не посещала.

3) Формирование *предсказаний* о результатах действий. Оценка *уверенности в предсказаниях*, стремление к надежным предсказаниям, уверенность в которых высока.

4) Формирование *базы знаний*, характеризующей ситуации, действия, прогнозы результатов действий, оценки расстояний между текущими ситуациями и целевой ситуацией.

5) Использование базы знаний при планировании целенаправленного поведения.

6) *Прямые и обратные мыслительные процессы* (от исходной ситуации к целевой и обратно).

7) Формирование *стереотипа* целенаправленного поведения.

Дополнительно мы рассмотрим возможность моделирования агентов, изобретающих аксиоматический подход.

#### **4. О формализации аксиоматического подхода**

Важный подход научного познания – аксиоматический, который возник в Древней Греции после появления математических доказательств и активного развития геометрии в школе Платона. В своих знаменитых «Началах» Евклид собрал и систематизировал важнейшие труды по математике, существовавшие в конце IV века до н.э. При возникновении аксиоматической теории характерно следующее:

- Подготовительный этап, для «Начал» Евклида этот этап включал в себя: 1) возникновение математических доказательств, 2) использование доказательств при установлении теорем, при решении многочисленных математических задач, 3) накопление знаний о полученных результатах и используемых методах.
- Формулировка основных понятий, введение определений.
- Формулировка основных положений.
- Формулировка аксиом.
- Использование понятий, положений, аксиом при развитии теории.

Отметим, что процессы возникновения доказательств и аксиоматического подхода хорошо охарактеризованы в книге В.Ф. Турчина «Феномен науки» (глава 10) [Turchin, 1977; Турчин, 1993].

«Начала» Евклида длительное время служили образцом построения теорий. Например, И. Ньютон в «Математических началах натуральной философии» явно использовал структуру «Начал» Евклида.

Может ли автономный агент использовать аксиоматический подход при расширении своей базы знаний?

В принципе, это можно себе представить, хотя процесс автоматического формирования аксиоматической теории автономным агентом может быть весьма непростым.

#### **Заключение**

Таким образом, в настоящей работе проанализированы различные аспекты проблемы: как автономный агент может прийти к самостоятельному познанию природы. Проведённый анализ показывает, что в обучении с



подкреплением явно проработаны методы, которые могут быть использованы при построении когнитивных агентов, познающих закономерности природы. Перечислим эти методы.

- 1) Предсказание будущих ситуаций, как независимых от действий агентов (см., например, [Редько, 2020]), так в результате таких действий.
- 2) Внутренние модели агентов.
- 3) Иерархия внутренних моделей [Панов, 2022; Dzhivelikian et al., 2022].
- 4) Внутренняя мотивация поведения агентов [Dzhivelikian et al., 2022].
- 5) Специальная потребность накопления знаний агента [Жданов, 2008; Непомнящих, 2012; Непомнящих, 2013; Непомнящих, 2014].
- 6) Возможность агента самостоятельно ставить цели своего поведения [Непомнящих, 2013; Непомнящих, 2014].
- 7) Использование ранее полученных знаний в текущем поведении (отложенное обучение) [Witkowski, 2007].
- 8) Построение модели психики агента [Шумский, 2019; Шумский, 2022].

Все эти методы могут быть использованы при создании когнитивных агентов, приближающихся к «агентам-учёным» [Red'ko, 2012; Schmidhuber, 2015].

## Список литературы

- [Жданов, 2008] Жданов А.А. Автономный искусственный интеллект. – М.: Бинном, 2008.
- [Зорина и др., 2001] Зорина З.А., Полетаева И.И. Зоопсихология. Элементарное мышление животных. – М.: АспектПресс, 2001.
- [Канеман, 2013] Канеман Д. Думай медленно... решай быстро. – М.: АСТ, 2013.
- [Непомнящих, 2012] Непомнящих В.А. Адаптация к решению частных задач и «глобальные» цели в поведении животных // Нейроинформатика (электронный журнал). – 2012. – Т. 6, № 1. – С. 12-22. – URL: <http://www.niisi.ru/iont/ni/Journal/V6/N1/Непомnyashchikh.pdf> (дата обращения: 16.08.2022).
- [Непомнящих, 2013] Непомнящих В.А. Адаптация и автономия в поведении животных // XV Всероссийская научно-техническая конференция «Нейроинформатика–2013»: Лекции по нейроинформатике. – М.: НИЯУ МИФИ, 2013. – С. 106-123.
- [Непомнящих, 2014] Непомнящих В.А. Связь между автономным и адаптивным поведением у искусственных агентов и животных // Подходы к моделированию мышления. – М.: УРСС, 2014. – С. 9-28.
- [Панов, 2022] Панов А.И. Иерархическая постановка задачи объектно-центричного обучения с подкреплением // Интегрированные модели и мягкие вычисления в искусственном интеллекте: Сборник научных трудов XI Международной научно-практической конференции ИММВ-2022. – М.: РАИИ, 2022. – Т. 2. – С. 248-256.

- [Редько, 2018] Редько В.Г. Моделирование когнитивной эволюции: На пути к теории эволюционного происхождения мышления. – 2-е изд., исп. и доп. – М. ЛЕНАНД/URSS, 2018.
- [Редько, 2020] Редько В.Г. Модель чувства причинности // Труды НИИСИ РАН. – 2020. – Т. 10, № 2. – С. 34-38. См. также: [https://trudy.niisi.ru/2020\\_T10\\_N2.pdf](https://trudy.niisi.ru/2020_T10_N2.pdf) (дата обращения: 16.08.2022).
- [Редько и др., 2021] Редько В.Г., Самсонович А.В. Формализация процессов озарения // Девятнадцатая Национальная конференция по искусственному интеллекту с международным участием КИИ-2021: Труды конференции. – Ростов-на-Дону; Таганрог: Изд-во ЮФУ, 2021. – С. 61-86.
- [Турчин, 1993] Турчин В.Ф. Феномен науки. Кибернетический подход к эволюции. – М.: Наука, 1993. – 1-е изд. – М.: ЭТС, 2000. – 2-е изд. См. также <http://www.refal.ru/turchin/phenomenon/> (дата обращения: 16.08.2022).
- [Ухтомский, 2002] Ухтомский А.А. Доминанта. Статьи разных лет. 1887-1939. – СПб.: Питер, 2002.
- [Шумский, 2019] Шумский С.А. Машинный интеллект. Очерки по теории машинного обучения и искусственного интеллекта. – М.: РИОР, 2019.
- [Шумский, 2022] Шумский С.А. ADAM – модель искусственной психики // Автоматика и телемеханика. – 2022. – № 6. – С. 24-37.
- [Юм, 1748/1966] Юм Д. Исследование о человеческом познании. Соч. в 2-х т. Т. 2. – М.: Мысль, 1966. – С. 5-169.
- [Dzhivelikian et al., 2022] Dzhivelikian E., Latyshev A., Kuderov P., Panov A. Hierarchical intrinsically motivated agent planning behavior with dreaming in grid environments // Brain Informatics. – 2022. – Vol. 9 (8). – URL: <https://doi.org/10.1186/s40708-022-00156-6> (date of access: 16.08.2022).
- [Inglis et al., 1997] Inglis I.R., Forkman B., Lazarus J. Free food or earned food: a review and fuzzy model of contrafreeloading // Animal Behaviour. – 1997. – Vol. 53, No. 6. – P. 1171-1191.
- [Inglis et al., 2001] Inglis I.R., Langton S., Forkman B., Lazarus J. An information primacy model of exploratory and foraging behaviour // Animal Behaviour. – 2001. – Vol. 62, No. 3. – P. 543-557.
- [Laukien, 2016] Laukien E., Crowder R., Byrne F. Feynman machine: The universal dynamical systems computer // ArXiv of Cornell University. – URL: <https://arxiv.org/abs/1609.03971> (date of access: 09.08.2022).
- [Reed et al., 2022] Reed S., Zolna K., Parisotto E. et al. A generalist agent // ArXiv of Cornell University. – URL: <https://arxiv.org/abs/2205.06175> (date of access: 04.08.2022).
- [Red'ko, 2012] Red'ko V.G. Principles of functioning of autonomous agent-physicist // Proc. of the Third Annual Meeting of the BICA Society. – Heidelberg, New York, Dordrecht, London: Springer, 2012. – P. 265-266.
- [Red'ko et al., 2015] Red'ko V.G., Nepomnyashchikh V.A., Osipova E.A. Models of fish exploratory behavior in mazes // Biologically Inspired Cognitive Architectures. – 2015. – Vol. 13. – P. 9-16.

- [Red'ko et al., 2016]** Red'ko V.G., Burtsev M.S. Modeling of mechanism of plan formation by New Caledonian crows // *Procedia Computer Science*. – 2016. – Vol. 88. – P. 403-408. – URL: <http://www.sciencedirect.com/science/article/pii/S1877050916317124> (date of access: 16.08.2022)
- [Schmidhuber, 2015]** Schmidhuber J. On learning to think: Algorithmic information theory for novel combinations of reinforcement learning controllers and recurrent neural world models // *ArXiv of Cornell University*. – URL: <https://arxiv.org/abs/1511.09249> (date of access: 07.08.2022).
- [Schmidhuber, 2018]** Schmidhuber J. One big net for everything // *ArXiv of Cornell University*. – URL: <https://arxiv.org/abs/1802.08864> (date of access: 07.08.2022).
- [Sutton et al., 2018]** Sutton R.S., Barto A.G. *Reinforcement Learning: An Introduction*. – Second Edition. – Cambridge, MA, MIT Press, 2018.
- [Turchin, 1997]** Turchin V.F. *The Phenomenon of Science. A Cybernetic Approach to Human Evolution*. – New York: Columbia University Press, 1977.
- [Witkowski, 2007]** Witkowski M. An action-selection calculus // *Adaptive Behavior*. – 2007. – Vol. 15, No. 1. – P. 73-97.

## ИЗВЛЕЧЕНИЕ АРГУМЕНТАЦИИ НА ОСНОВЕ ИНДИКАТОРНОГО ПОДХОДА

Е.А. Сидорова (*lsidorova@iis.nsk.su*)  
И.Р. Ахмадеева (*i.r.akhmadeeva@iis.nsk.su*)  
И.С. Кононенко (*irina\_k@cn.ru*)  
П.М. Чагина (*p.chagina@gmail.com*)  
Институт систем информатики  
им. А.П. Ершова СО РАН, Новосибирск

В статье рассматривается индикаторный подход к извлечению аргументации, встречающейся в научно-популярной литературе. Представлены типы индикаторов аргументации, их соотношение с множеством дискурсивных маркеров и приведена методика составления словаря индикаторов аргументации. Предложен подход к использованию индикаторов аргументации в методах глубокого обучения, основанный на анализе контекста использования индикаторов. Для построения обучающей выборки для каждого индикатора извлекаются основное утверждение, а также левый и правый контекст, как правило, представленные соседними предложениями. Каждый набор размечается по наличию аргументации в соответствии с разметкой корпуса. Для построения классификатора использовался список из 143 индикаторов аргументации и размеченный корпус, включающий 162 статьи научно-популярного жанра, размещенных на веб-платформе ArgNetBank Studio. Всего на основе корпуса было получено около 4600 обучающих контекстов. Результаты приведенных экспериментов по извлечению аргументации показали лучшее качество работы классификатора, основанного на индикаторах.

**Ключевые слова:** аргументация, извлечение аргументации, схема аргумента, индикаторы аргументации, индикаторный подход.

### Введение

Извлечение и анализ аргументации (Argument Mining) – современное направление исследований, относящееся к области компьютерной лингвистики, задачей которого является автоматическое извлечение из текстов множества логически связанных утверждений или аргументативных цепочек.

чек, являющихся обоснованием или опровержением некоторого базового тезиса. Данная область активно развивается в последнее десятилетие во всем мире и уже имеет множество приложений, таких как анализ политических дебатов, прогнозирование на основе анализа высказываний пользователей, оценка аргументированности научных статей, анализ юридических документов. Данная задача имеет также потенциальную возможность применения в области анализа фейков и в вопросно-ответных системах для организации диалога или рассуждений на основе фактического материала.

В России данное направление развивается последние 3 года. В частности, первый русскоязычный корпус с разметкой аргументации был создан нашим коллективом в 2019-20 гг. Для создания данного ресурса была разработана отечественная система разметки корпусов ArgNetBank Studio [Сидорова и др., 2020] на основе известной системы OVA+ (Online Visualisation of Argument) [Janier et al., 2014] и онтологии аргументации. Чуть позже коллективом российских ученых был создан переводной корпус с разметкой [Fishcheva et al., 2019; Fishcheva et al., 2021] на базе англоязычных ресурсов: корпуса аргументативных микротекстов (ArgMicro) [Peldszus et al., 2015] и корпуса убедительных эссе (PersEssays) [Stab et al., 2014]. Полученный корпус имеет более примитивную модель разметки аргументации, в соответствии с используемыми источниками. В 2022 г. на конференции Диалог было проведено соревнование по поиску аргументирующих высказываний на заданную тему, для чего организаторами был создан тематический корпус, содержащий комментарии к постам в социальных медиа, связанным с пандемией COVID-19 [Kotelnikov et al., 2022].

Способы аргументации активно исследовались еще в середине прошлого столетия в рамках философских наук. Результатом данных исследований явилось создание модели аргументации С. Тулмина [Toulmin, 2003], определяющей состав и формальную структуру аргумента и его связей, а также разработка открытого набора типовых схем аргументации Д. Уолтона [Walton et al., 2008], которые на текущий момент являются де-факто стандартом в области аргументации (по крайней мере, при исследовании аргументации с помощью информационных технологий). На основе данного набора была создана онтология аргументации (AIF-онтологии) [Rahwan et al., 2011], что позволило применить технологии Semantic Web для унификации разметки, организации хранилища и создания поисковых сервисов. Следует отметить, что набор является открытым, поскольку каждый функциональный жанр текста накладывает свои ограничения на набор используемых схем рассуждений.

Филологами давно отмечена роль дискурсивных коннекторов как ключевых индикаторов структуры дискурса, что нашло отражение в аннотации Penn Discourse Treebank (PDTB), которая охватывает традиционные

функциональные слова и фразы, такие как подчинительные союзы (например, *when, because, as soon as*), сочинительные союзы (*and, but, or*), наречия (*instead, therefore*), предложные группы (*on the other hand*) и т.д. [Prasad et al., 2007]. Индикаторы того или иного аспекта содержания текста представляют собой специфические подсказки не только в виде функциональных слов, но и в виде различного рода словесных клише (образцов, маркеров), в общем случае разрывных: так для текста научной статьи характерны индикаторы *"в настоящей работе"*, *"в работе... рассматривается"*, *"целью...является"*. Для выявления фрагментов, содержащих аргументацию, и определения достоверности и убедительности аргументов существенную роль играют показатели аргументации, наличие которых доказано в теоретическом исследовании [Van Eemeren et al. 2007], где демонстрируется возможность рассмотрения широкого класса языковых выражений, сигнализирующих о применении той или иной схемы аргументации, причем множество этих выражений выходит далеко за рамки функциональных классов слов: оно включает также конструкции со знаменательными словами (ср. русскоязычные *"отсюда следует"*, *что*, *"считается, что"*, *"на этом основании... можно предположить, что"*, *"аналогом этому служит"*, и т.п.). Поэтому при разработке платформы для анализа аргументации мы включили в модель разметки маркеры, которые могут, по мнению экспертов, играть важную роль при выявлении и оценке аргументации, а также снабдили систему сервисами по автоматическому извлечению маркеров из корпусов текстов и построению конкордансов [Кононенко и др., 2020].

Целью нашего текущего исследования является дальнейшее изучение роли индикаторов в обнаружении и типизации аргументации, а также способов их выявления и оценки эффективности. С этой целью мы сформировали словарь индикаторов на основе существующих словарей дискурсивных маркеров, предложили модель для обучения нейронной сети на основе индикаторного анализа текста и провели набор экспериментов по выявлению аргументов. Дальнейшая часть статьи организована следующим образом. В первой главе рассмотрены подходы, наиболее актуальные для области извлечения аргументации. Вторая глава посвящена вопросам создания словаря индикаторов. В третьей главе описывается подход к использованию индикаторов для применения в методах глубокого обучения, рассматривается вопрос подготовки данных. И, наконец, в четвертой главе представлены результаты экспериментов и проведен их анализ.

## 1. Обзор работ по извлечению аргументации

Автоматическое выявление аргументов предполагает решение следующих задач:

- 1) выявление аргументативных/неаргументативных зон и сегментация текста с выделением релевантных для аргументации дискурсивных единиц (ADU);
- 2) выделение компонентов аргумента и установление их ролей в составе аргумента (посылка, заключение);
- 3) установление отношений и их типов (атака, поддержка, конфликт) между аргументами, построение базовой аргументативной структуры;
- 4) уточнение выделенных структур аргументов путем сопоставления им схем аргументации, представляющих типовые модели рассуждения и убеждения.

На каждом этапе анализа может быть использован подход как на основе методов машинного обучения, так и на базе лингвистических знаний в виде правил, составленных экспертом, либо комбинация этих подходов. При наличии специальным образом размеченных корпусов текстов, преимущество имеют методы машинного обучения, не требующие больших временных затрат на детальное изучение объекта поиска. Выбор метода машинного обучения определяется экспериментально, т.к. очевидных предпосылок для предпочтения того или иного алгоритма нет [Lawrence et al., 2019]. Показано, что качество поиска методами машинного обучения существенно зависит прежде всего от выбора признаков, а также от сочетания методов машинного обучения с методами на основе правил.

Исследователи, решающие задачу разделения утверждений на аргументативные и неаргументативные, нередко проводят анализ на уровне отдельных предложений, а не клауз. Аргументативные предложения как поисковые объекты фигурируют в работах [Levy et al., 2014; Sardianos et al., 2015]. Часть работ посвящается обнаружению аргументативных предложений \ клауз, играющих роль заключения и/или посылки, например, [Nitin et al., 2012]. Для поиска аргументативных предложений (или клауз) традиционно применяют такие методы машинного обучения как: PNB и максимум энтропии, SVM и логистическая регрессия, деревья решений, нейронные сети.

Используемые системы признаков, как правило, включают несколько типов характеристик: структурные, лексические, синтаксические, контекстные, индикаторы. В качестве признаков рассматриваются n-граммы ( $n \geq 1$ ), пары слов, теги отдельных частей речи, грамматические характеристики (время, лицо глагола и др.), число знаков препинания [Goudas et al., 2014], позиции в предложении (основная или вспомогательная, например, в при-

даточном предложении), дискурсивные маркеры [Park et al., 2014] [Stab et al., 2018]. В дополнение к перечисленным признакам используют информацию о типах поисковых объектов (аргументов или контраргументов, посылок или заключений), о риторической структуре [Park et al., 2014], тематической структуре текста [Hua et al., 2017] и пр. Наилучшие комбинации признаков и методов машинного обучения, как правило, выбираются экспериментально. Обычно используют бинарные значения признаков.

Исследование аргументации в юридических текстах в [Sardianos et al., 2015] показало, что наилучшей комбинацией "система признаков + метод" является алгоритм PNB и максимум энтропии, обученные на  $n$ -граммах ( $n = 1, 2, 3$ ), отдельных частях речи, ключевых словах, пунктуации и текстовой статистике. Точность выявления аргументативных предложений составила около  $P = 0.73$ . В работе [Hua et al., 2017] показано, что использование контекста слов при обучении нейронной сети, а также структурных, синтаксических, лексических и парадигматических признаков позволяет определить границы аргументов с  $F = 0.89$ .

Что касается анализа аргументации для русского языка, то первые попытки таких исследований и создания необходимых наборов данных предприняты совсем недавно. Так, в 2022 году в рамках конференции «Диалог» проводилось первое соревнование русскоязычных систем анализа аргументации RuARG-2022 [Kotelnikov et al., 2022]. В ходе соревнования участникам были предложены две задачи: определение позиции автора по заданной теме (аргументативность) и классификация доводов («за» или «против»). Лучшие результаты были получены на основе моделей (ru)BERT, (crosslingual)RoBERTa(-large). Оценки F-меры составили: а) без использования дополнительных данных  $F1 = 0.65$  и  $0.7$ , соответственно, б) при использовании дополнительных данных  $F1 = 0.69$  и  $0.74$ , соответственно. Для получения дополнительных данных использовались методы генерации парафраз mT5, дополнительные предметно-ориентированные наборы данных на английском языке (с этой целью базовый набор данных автоматически переводился на английский язык), случайное сэмплирование и др. В целом было показано, что любые дополнительные данные полезны для решения поставленных задач.

## 2. Индикаторы аргументации

В работе [Kononenko et al., 2019] демонстрируется возможность типизации индикаторов с различных точек зрения: прагматический аспект аргументации, степень грамматикализации, семантика опорного слова индикатора, тип конструкции. Так, показано, что индикатор может сигнализировать о таких прагматических аспектах аргументации, как:

- мнение и сила аргумента (степень уверенности);



- отношение вывода между двумя утверждениями;
- роль утверждения в отношении вывода (Посылка vs. Заключение);
- тип аргументативного отношения (Поддержка vs. Атака);
- структура аргументации (Множественная vs. Последовательная аргументация);
- семантико-онтологическое отношение, на котором основана применяемая в данном случае типовая схема рассуждения;

Аналогично выводам работы [Toldova et al., 2018] в отношении показателей риторических отношений, рассматривается деление индикаторов аргументации на первичные (зафиксированные в грамматиках и словарях функциональные слова) и вторичные (построенные на основе полнозначных лексем):

- дискурсивные коннекторы – известные функциональные единицы, в том числе многословные (предлоги, союзы, вводные слова): поэтому, поскольку, следовательно, так как, значит, тем более что, например, в частности и т. п.;
- знаменательные слова и конструкции на их основе, которые, в свою очередь, классифицируются по семантике опорного знаменательного слова (предикаты вывода, речевые и ментальные предикаты и т. п.).
- На базе речевых и ментальных предикатов, предикатов вывода и ментального воздействия формируются сложные показатели аргументации. Помимо основного слова, они могут включать маркеры позиций сентенциальных актантов, например, союз что, местоименно-соотносительную конструкцию то, что, анафорические и катафорические элементы, такие как указательное местоимение это/этот, наречие отсюда, относительное местоимение что.

В работе [Ахмадеева и др., 2019] показано, что деятельность эксперта по наполнению словаря индикаторов аргументации включает такие этапы, как выделение в коллекции текстов слов и словосочетаний – кандидатов на роль индикаторов; контекстный анализ индикаторов для представления индикаторных структур в виде шаблонов; обобщение лексического состава шаблонов до лексико-семантического уровня. Для данного эксперимента было необходимо обнаружить в текстах коллекции дискурсивные слова и конструкции и отобрать те, которые могут претендовать на роль маркеров аргументации. Для того, чтобы уменьшить объем ручной работы, в настоящем эксперименте предварительный этап отбора был сокращен, что стало возможным благодаря двум факторам. С одной стороны, учтена гипотеза о том, что многие показатели аргументации являются подмножеством всего множества дискурсивных маркеров (либо существенно пересекаются с ним). Эта гипотеза основана на противопоставлении предметных и презентационных отношений в Теории риторической структуры [Mann et al., 1988]. Целью презентационных риторических от-

ношений является усиление предрасположенности читателя, например, желания действовать или позитивного отношения, уверенности или принятия утверждения-ядра риторического отношения. С этой точки зрения презентационные риторические отношения пересекаются с аргументативными дискурсивными отношениями. Отображение риторических отношений на аргументативные структуры, выполненное в работе [Stede et al., 2016] подтверждает это прагматическое сходство. Не удивительно, что в первых экспериментах по извлечению аргументации в качестве лексических индикаторов используются традиционные функциональные слова. С другой стороны, в открытом доступе имеется современный разножанровый русскоязычный корпус Ru-RSTreebank (<https://rstreebank.ru>) [Pisarevskaya et al., 2017], в котором есть выделенный при аннотировании риторических структур список дискурсивных маркеров. Этот список и был принят в качестве исходного.

Все элементы исходного списка были автоматически трансформированы в формальное представление на языке шаблонов, позволяющем генерировать служебные шаблоны (которые могут содержать альтернативы) и указывать наличие разрывов:

*// тот факт, что ... говорит о том,*

**[говорит о]** = [begin!: 'говорить', end!: 'о', not\_include: [**пunct**]]

**[том]** = ['том'] ['тома']

**[говорит о том]** = [ [**говорит о**], [**том**], ', ' ]

**[тот факт что]** = ['том', 'факт', ';', 'что']

**[факт говорит о]** = [begin!: [**тот факт что**], end!: [**говорит о том**], not\_include: [**пunct**]]

Это представление используется поисковой системой при указании соответствующего маркера и подкорпусов в качестве параметров поиска.

Работа эксперта заключалась в анализе исходного списка маркеров и исключении тех единиц, которые не могут претендовать на роль индикаторов аргументации. Работа проводилась с использованием встроенного в платформу корпус-менеджера, позволяющего подсчитывать частоты единиц в корпусе и подкорпусах, находить и анализировать все контексты употребления маркера, вплоть до целого текста.

Основанием для исключения маркера из списка служили следующие факторы:

- маркеры с нулевой частотой в корпусе;
- редкие маркеры с низкой частотой в корпусе;

- неаргументативные маркеры, не встретившиеся ни в одном аргументативном контексте;
- неоднозначные маркеры, которые в равной степени или реже обнаруживаются в аргументативном контексте, чем неаргументативном;
- вложенные маркеры, покрытые шаблонами других маркеров, уже включенных в словарь индикаторов аргументации.

В табл. 1 приведены примеры дискурсивных маркеров, исключенных из словаря индикаторов аргументации.

Таблица 1

### Примеры исключаемых дискурсивных маркеров

Типы маркеров	Примеры маркеров	Примеры контекстов
с нулевой частотой	<i>в противном случае притом X обуславливать Y связан скорее с тем, что ... чем с</i>	
редкие	X <i>нельзя сказать о</i> (6)	Автомобили, самолеты, средства телекоммуникации резко сократили субъективные масштабы нашей планеты и практически превратили ее в единую систему, в которой все части сильно взаимосвязаны. Этого <b>нельзя сказать о</b> глобализации духовной стороны жизни человечества.
неаргументативные	<i>либо наоборот</i> (2) <i>в условиях</i> (416) <i>при этом</i> (2453) <i>причем</i> (1347)	Такая система может находиться в одном из двух состояний: либо ген А «работает», и тогда ген Б «выключен», <b>либо наоборот</b> . <представление альтернатив>
неоднозначные	<i>пусть</i> (447)	<b>Пусть</b> все больше детей приезжает на олимпиады, пусть будет больше любителей русского языка, больше единомышленников! <побуждение> Так, например, <b>пусть</b> надежность работы каждого элемента (сегмента) равна 0,9999. <условное допущение> А нам надо продолжать делать свое дело, <b>пусть</b> это и выглядит порой как Сизифов труд, <уступительное допущение – <i>arg</i> >

	<i>X приводит в Y (38)</i>	<b>приводит в</b> замешательство/ ярость... <каузация состояния> <b>приводил в</b> лабораторию <каузация перемещения> <b>приводит в</b> статью/в пример <цитирование – арг>
вложенные	<i>признать, (что) (152)</i>	Специалисты открыто <b>признают, что</b> сосудистая хирургия всех проблем в этой области не решает <экспертное мнение – арг>

Исходный список маркеров содержал 294 единицы, в результате анализа была исключена 151 единица. Итоговый список кандидатов состоит из 143 маркеров.

### 3. Извлечение аргументации на основе индикаторов

Задача анализа аргументации заключается в автоматическом извлечении из текстов множества высказываний (посылок), приводящих к некоторому заключению (утверждению). Определить аргумент в тексте часто можно по наличию индикатора аргументации. В таком случае поиск аргументативных предложений можно начать именно с поиска маркеров и анализа контекста их использования. Для каждого индикатора контекст его использования разделяется на три части: основное утверждение, включающее индикатор, а также правый и левый контексты. При этом некоторые индикаторы, указывающие на наличие логической связи между посылкой и выводом аргумента, позволяют однозначно определить, где находится посылка, а где заключение. Рассмотрим примеры таких индикаторов.

(1) *С другой стороны, понятно, что без нормы мы можем оказаться в пространстве социального непонимания, **потому что** речевая коммуникация является основой социальной коммуникации, и здесь нужны «правила игры». <заключение до, посылка после>.*

(2) *Логика и созвучность слов склоняют Вас к мнению о том, что кириллица – это вне сомнений та азбука, что изобретена Кириллом. **Тем не менее**, дошедшие до нас старославянские источники не дают однозначной информации. <заключение после, посылка до>.*

Группируя такие потенциально аргументативные высказывания в соответствии с положением относительно индикатора, мы объединяем не просто аргументативные предложения, а конкретно посылки и выводы, что может увеличить точность автоматического распознавания аргументов, обученного на такой модели.

Для получения обучающего массива данных использовались 2 корпуса, созданные с помощью веб-платформы “Программный комплекс для моделирования и анализа аргументации в научно-популярных текстах ArgNetBank Studio” (<https://uniserv.iis.nsk.su/arg> и <https://geos.iis.nsk.su/arg>). Первый корпус был создан для филологического исследования научно-популярного жанра, второй - для исследования связи между риторической и аргументативной разметкой на базе научно-популярных новостных сообщений. Аргументативная разметка корпусов осуществлялась на основе расширенной онтологии AIF (Argument Interchange Format), построенной в соответствии с теорией Д. Уолтона. Полученный корпус состоит из 162 текстов, содержащих разметку, при этом не учитывались тексты с малым объемом разметки (менее 5 аргументов), а при наличии нескольких экземпляров разметки одного текста выбирался максимальный по объему.

Для обработки корпуса использовался список из 143 индикаторов аргументации. Для построения обучающей выборки для каждого найденного в корпусе индикатора извлекались основное утверждение, а также левый и правый контекст, которые, как правило, были представлены соседними предложениями. Каждый набор размечается по наличию аргументации в соответствии с разметкой корпуса.

Данные для обучения представляются в следующем виде:

pattern | left\_context | main\_st | right\_context | left\_arg | main\_arg | right\_arg,

где

- имя индикатора аргументации (pattern) – название найденного индикатора;
- основное утверждение (main\_st) – предложение, содержащее индикатор; для выделения индикатора в тексте утверждения использовался стандартный способ маркирования с помощью специального пунктуационного знака (\*);
- левый контекст (left\_context) – часть предложения, предшествующая вхождению индикатора, объединенная с предложением, предшествующим основному утверждению (если оно имеется);
- правый (right\_context) контекст – часть предложения после индикатора, объединенная с предложением, следующим за основным утверждением (если оно имеется);
- левый (left\_arg), внутренний (main\_arg) и правый аргументы (right\_arg) – числовые значения наличия аргумента в соответствующем контексте, принимают значения 0 и 1.

Приведенному выше примеру (2) будет соответствовать единица данных:

pattern = *тем не менее*  
left\_context = “*Логика и созвучность слов склонят Вас к мнению о том, что кириллица – это вне сомнений та азбука, что изобретена Кириллом.*”  
main\_st = “*\*Тем не менее\*, дошедшие до нас старославянские источники не дают однозначной информации.*”  
right\_context = “*дошедшие до нас старославянские источники не дают однозначной информации.*”  
left\_arg = 1 | inner\_arg = 1 | right\_arg = 1

Таким образом, на основе индикаторов аргументации из корпуса было извлечено около 4600 строк такого вида. Выборка была разделена на обучающую, тестовую и валидационную части. Для контроля целостности разделение осуществлялось на уровне целых текстов (а не на данных). Для повышения точности оценки экспериментов из корпуса в тестовую выборку были отобраны 34 текста (около 20% корпуса) с наиболее качественной разметкой (разметка была проверена и согласована двумя экспертами).

#### 4. Экспериментальное исследование

Для представления русского текста использовалась модель ruBERT (DeepPavlov/rubert-base-cased-conversational), на основе которой получались эмбединги предложений. Мы использовали следующие конфигурации, основанные на индикаторном подходе выбора контекста:

- 1) независимая классификация (simple-model): классификатор применяется к эмбедингам предложений, объединяющих основное утверждение, левый и правый контексты (три утверждения объединяются в одно и векторизуются);
- 2) классификация с учетом индикаторов (indicator-model): отличие от simple-model заключается в том, что дополнительно в основном утверждении размечаются индикаторы аргументации с помощью специального знака пунктуации (\*).

В обоих случаях классификаторы предсказывали наличие аргументации в основном утверждении.

Классификаторы включают модель BERT и полносвязную нейронную сеть. Для всех описанных выше конфигураций использовались одни и те же параметры: коэффициент скорости обучения = 0.00002, размер пакета (batch size) = 16, количество эпох = 10. Полносвязная сеть состоит из линейного слоя размера (768, 1024), активационной функции ReLU, слоя исключения (dropout) с вероятностью 0,6, еще одного линейного слоя размера (1024, 2), слоя softmax. Такие параметры классификатора как

dropout, размер скрытого слоя и коэффициент скорости обучения подбирались в зависимости от качества работы классификатора на валидационной выборке.

Оценка качества работы классификаторов была проведена на двух различных текстовых выборках. Для первого эксперимента были взяты тексты, выборка на которых давала примерно одинаковое распределение между положительными и отрицательными примерами. Для второго эксперимента в тестовую выборку отбирались тексты с проверенным качеством разметки.

Таблица 2

### Результаты экспериментов по извлечению аргументов

Тип классификатора	Precision	Recall	F1-мера
simple-model(1)	59,07	84,58	69,56
indicator-model(1)	59,08	86,79	<b>70,3</b>
simple-model(2)	82,13	78,91	80,49
indicator-model(2)	<b>83,26</b>	<b>83,81</b>	<b>83,53</b>

Хорошие показатели, демонстрируемые simple-model(2) и indicator-model(2), связаны с тем, что для тестирования использовались только тексты, размеченные экспертами. Такая разметка достаточно полно покрывает текст, что в итоге привело к тому, что 78,7% предложений, вошедших в тестовую выборку, оказались аргументативными (vs. 20,7% для обучающей выборки).

В целом, данные экспериментов показывают, что на данном корпусе использование индикаторов улучшает качество работы классификатора по всем трем показателям. В дальнейшем мы планируем изучить роль индикаторов и на других корпусах.

### Заключение

В данной статье представлен оригинальный подход к исследованию роли индикаторов при извлечении аргументации. Результаты первого этапа исследований показали качество предложенного нами метода извлечения аргументов на основе индикаторов из тематически нейтральных русскоязычных текстов, сопоставимое с современными нейросетевыми подходами, представленными в [Kotelnikov et al., 2022]. Также полученные результаты согласуются с наилучшими показателями предыдущих экспериментов на данном корпусе (67,8% F-меры), приведенных в работе [Salomatina et al., 2021]. Однако проводить сравнение с данными показателями можно лишь приблизительно, поскольку в текущем эксперименте оценка проводилась не на всем корпусе предложений, а только на предложениях с индикаторами.

Следует отметить сложность адекватной оценки качества работы анализаторов, связанную в первую очередь с неполнотой разметки корпуса и проблемой согласованности между аннотаторами (для неопытных аннотаторов  $k = 0.58$ , а для экспертов  $k = 0.83$  [Lawrence et al., 2019]). Также существенную роль играет качество словаря индикаторов, полученного из более общего словаря дискурсивных маркеров. Таким образом, дальнейшие исследования будут связаны с изучением следующих вопросов:

- пополнение словаря индикаторов аргументации, в частности, планируется привлечь словари синонимов и переводные словари;
- увеличение объемов обучающих выборок за счет автоматического перевода имеющихся иноязычных корпусов с разметкой аргументации на основе теории Д. Уолтона, а также использование корпусов с риторической разметкой;
- улучшение качества аннотирования за счет выработки методологии разметки на базе теории Д. Уолтона;
- расширение экспериментальной базы и внедрение классификации в соответствии со схемами аргументации, актуальными для заданного жанра.

Важный вопрос в контексте применения методов на основе индикаторов: “Как обнаружить аргументы, в которых не используются явные индикаторы?” По нашим наблюдениям, объем таких аргументов составляет около 20%. Мы считаем, что такие аргументы будет легче обнаружить, опираясь на уже найденные аргументы с индикаторами, используя методы определения тематической схожести и анализ связности аргументации [Саломатина и др., 2020].

### Список литературы

- [Ахмадеева и др., 2019] Ахмадеева И.Р., Кононенко И.С., Саломатина Н.В., Сидорова Е.А. Подход к построению шаблонов индикаторов для извлечения аргументов из научно-популярных текстов // Тр. Международной конференции “Знания – Онтологии – Теории” (ЗОНТ-2019). – Институт математики им. С.Л. Соболева СО РАН, Новосибирский государственный университет, 2019.
- [Кононенко и др., 2020] Кононенко И.С., Ахмадеева И.Р., Сидорова Е.А. Лингвистические аспекты исследования аргументации на основе онтологии // Информационные и математические технологии в науке и управлении. – 2020. – № 4 (20).
- [Саломатина и др., 2020] Саломатина Н.В., Кононенко И.С., Сидорова Е.А., Пименов И.С. Распознавание аргументативных связей в научно-популярных текстах // Системная информатика. – 2020. – № 16.
- [Сидорова и др., 2020] Сидорова Е.А., Ахмадеева И.Р., Загорюлько Ю.А., Серый А.С., Шестаков В.К. Платформа для исследования аргументации в научно-популярном дискурсе // Онтология проектирования. – 2020. – Т. 10, № 4 (38).



- [Fishcheva et al., 2019] Fishcheva I., Kotelnikov E. Cross-Lingual Argumentation Mining for Russian Texts // Analysis of Images, Social Networks and Texts. – 2019.
- [Fishcheva et al., 2021] Fishcheva I., Goloviznina V., Kotelnikov E. Traditional machine learning and deep learning models for argumentation mining in russian texts // In Computational Linguistics and Intellectual Technologies: Proc. of the International Conference “Dialog-2021”. – 2021.
- [Goudas et al., 2014] Goudas T., Louizos C., Petasis G., Karkaletsis V. Argument extraction from news, blogs, and social media // In Artificial Intelligence: Methods and Applications. Vol. 8445. Lecture Notes in Computer Science. – Springer, 2014.
- [Hua et al., 2017] Hua X., Wang L. Understanding and Detecting Supporting Arguments of Diverse Types // Proc. of the 55th Annual Meeting of the Association for Computational Linguistics. Vol. 2: – Short Papers. 2017.
- [Janier et al., 2014] Janier M., Lawrence J., Reed C. OVA+: An argument analysis interface // Computational Models of Argument: Proceedings of COMMA. – 2014. – Vol. 266, No. 2014.
- [Kononenko et al., 2019] Kononenko I., Sidorova E. Development of the Lexicon of Argumentation Indicators / In: Kuznetsov S., Panov A. (eds) Artificial Intelligence. RCAI 2019. Communications in Computer and Information Science. Vol. 1093. – Springer, Cham, 2019.
- [Kotelnikov et al., 2022] Kotelnikov E., Loukachevitch N., Nikishina I., Panchenko A. RuArg-2022: Argument Mining Evaluation // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog-2022”. – 2022.
- [Lawrence et al., 2019] Lawrence J., Reed C. Argument mining: A survey. Int. J. of Computational Linguistics. – 2019. – No. 45(4). – P. 765-818.
- [Levy et al., 2014] Levy R., Bilu Y., Hershcovich D., Aharoni E., Slonim N. Context Dependent Claim Detection // Proc. of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, 2014.
- [Mann et al., 1988] Mann W.C., Thompson S.A. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization // Text. – 1988. – No. 8 (3).
- [Nitin et al., 2012] Nitin M., Heilman M., Tetreault J., Chodorow M. Identifying high-level organizational elements in argumentative discourse // Proc. of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2012. Montreal.
- [Park et al., 2014] Joonsuk Park, Claire Cardie. Identifying Appropriate Support for Propositions in Online User Comments // In Proceedings of the First Workshop on Argumentation Mining 2014. – P. 29-38.
- [Peldszus et al., 2015] Peldszus A., Stede M. An annotated corpus of argumentative microtexts // Argumentation and Reasoned Action: Proc. of the 1st European Conference on Argumentation. – London: College Publications; – 2015. – No. 2. – P. 801-815.
- [Pisarevskaya et al., 2017] Pisarevskaya D., Ananyeva M., Kobozeva M., Nasedkin A., Nikiforova S., Pavlova I., Shelepov A. Towards building a discourse-annotated corpus of Russian // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2017”. – 2017. – P. 194-204.

- [Prasad et al., 2007] Prasad R., Miltsakaki E., Dinesh N., Lee A., Joshi A., Robaldo L., Webber B. The Penn Discourse Treebank 2.0 Annotation Manual // Technical Report 203, Institute for Research in Cognitive Science, University of Pennsylvania. 2007.
- [Rahwan et al., 2011] Rahwan I, Banihashemi B, Reed C, Walton D, Abdallah S. Representing and classifying arguments on the semantic web // The Knowledge Engineering Review. – 2011. – No. 26 (4).
- [Salomatina et al., 2021] Salomatina N.V., Pimenov I.S., Sidorova E.A. Identification of argumentative sentences in Russian scientific and popular science texts // International Conference «Marchuk Scientific Readings 2021» (MSR-2021). – Novosibirsk, Russian Federation. Journal of Physics: Conference Series. – 2021. – Vol. 2099.
- [Sardianos et al., 2015] Sardianos C., Katakis I.M., Petasis G., Karkaletsis V. Argument Extraction from News // Proc. of the 2nd Workshop on Argumentation Mining. 2015.
- [Stab et al., 2014] Stab C., Gurevych I. Annotating argument components and relations in persuasive essays // Proc. of the International Conference on Computational Linguistics. – 2014.
- [Stab et al., 2018] Stab C., Miller T., Schiller B., Rai P., Gurevych I. Cross-topic argument mining from heterogeneous sources // Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018). – 2018.
- [Stede et al., 2016] Stede M., Afantenos S.D., Peldszus A., Asher N., and Perret J. Parallel discourse annotations on a corpus of short texts // Proc. of the International Conference on Language Resources and Evaluation (LREC), Portoroz. 2016.
- [Toldova et al., 2018] Toldova S., Pisarevskaya D., Vasilyeva M., Kobozeva M. The cues for rhetorical relations in Russian: "Cause-Effect" relation in Russian Rhetorical Structure Treebank // Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue". – 2018. – Issue 17.
- [Toulmin, 2003] Toulmin S. The Uses of Argument // Cambridge: Cambridge University Press, 2003.
- [Van Eemeren et al., 2007] Van Eemeren F.H., Houtlosser P., and F. Snoeck Henkemans. Argumentative Indicators in Discourse: A Pragma-Dialectical Study // Dordrecht: Springer. 2007.
- [Walton et al., 2008] Walton D., Reed C., Macagno F. Argumentation schemes. – Cambridge: Cambridge University Press. 2008.
- [Xie et al., 2020] Xie Q., Dai Z., Hovy E., Luong T., Le Q.V. Unsupervised data augmentation for consistency training // Advances in Neural Information Processing Systems. – 2020. – Vol. 33.

## ПОРОЖДЕНИЕ ИНТЕРПРЕТИРУЕМЫХ ВЕКТОРНЫХ ПРЕДСТАВЛЕНИЙ СЛОВ НА ОСНОВЕ СУПЕРПОНЯТИЙ\*

М.М. Тихомиров (*tikhomirov.mm@gmail.com*)

Н.В. Лукашевич (*louk\_nat@mail.ru*)

МГУ имени М.В. Ломоносова, Москва,

Центр доверенного искусственного интеллекта

ИСП РАН, Москва

В данной работе представлен подход по созданию интерпретируемых векторных представлений слов, в которых каждая компонента вектора соответствует некоторой интерпретируемой семантической категории. Для получения таких категорий используется лексико-семантический ресурс в виде семантической сети RuWordNet, а также представительный корпус русскоязычных текстов для порождения векторных представлений. Полученные интерпретируемые векторные представления были протестированы на возможность отображать разные модели в одно векторное пространство.

**Ключевые слова:** векторные представления слов, семантические категории, тезаурус, семантическая близость.

### Введение

В настоящее время основной формой представления слов и других языковых единиц являются представления в виде векторов низкой размерности (эмбедингов), получаемых за счет обработки больших текстовых корпусов [Mikolov et al., 2013b; Devlin et al., 2019]. При этом сами размерности векторов не являются интерпретируемыми – невозможно сказать, что означает та или иная размерность в векторе. Векторные представления формируются при анализе контекстов в корпусе, при этом в процессе обучения присутствует элемент случайности, в первую очередь при инициализации модели. Повторное вычисление векторных представлений на том же корпусе, векторные представления, рассчитанные на дру-

---

\* Работа поддержана грантом исследовательских центров по искусственному интеллекту ((соглашение 000000D730321P5Q0002 No. 70-2021-00142 с ИСП РАН).

гом корпусе, не могут быть сопоставлены между собой, требуют специальных процедур построения сопоставлений (преобразований) между векторными представлениями.

Проблема сопоставления разных векторных представлений, посчитанных на разных корпусах одного языка или разных языков, возникает в таких задачах, как кросс-языковые переносы моделей [Artetxe et al., 2009; Mikolov et al., 2013a], переносы моделей между различными областями (domain transfer) [Yamane et al. 2016], сравнение векторных представлений, обученных на текстовых корпусах, относящихся к различным временным интервалам (например, для изучения изменения значений слов).

В экспериментах разных авторов было показано, что векторные пространства содержат так называемые хабы (hubs), которые являются векторами, близкими к большому количеству других векторов в пространстве [Radovanovic et al., 2010]. Эта проблема проявляется в векторных пространствах слов в виде слов, которые имеют высокое косинусное сходство с большим количеством других слов [Ruder et al., 2019].

В данном исследовании мы рассматриваем проблемы хабов с другой стороны – можно ли выделить семантически интерпретируемые хабы в векторном пространстве, как некоторые существенные семантические категории в коллекции текстов, и далее развернуть векторное пространство с учетом этих хабов, сделав таким образом измерения векторного пространства интерпретируемыми.

Мы исследуем подход получения семантически интерпретируемых категорий для векторных представлений на основе привлечения лексико-семантических ресурсов в виде семантических сетей типа WordNet [Fellbaum, 1998]. В данном исследовании мы проводим эксперименты для русского языка, и в качестве таких хабов мы используем семантические классы (суперпонятия) из семантического ресурса русского языка RuWordNet. [Loukachevitch et al., 2016]. Отбор суперпонятий производится на основе представительного текстового корпуса, что позволяет создать адаптивный набор семантических категорий (суперпонятий) для задачи и/или предметной области. Полученные интерпретируемые векторные представления тестируются в задаче предсказания семантической близости слов.

## 1. Близкие работы

В теоретической и компьютерной семантике есть понятие семантического класса, т.е. совокупности слов, которые обладают похожими семантическими признаками, например *живые существа*, *предметы*, *действия* и др. Разные семантические теории и модели предлагают свои наборы семантических классов [Апресян, 1974; Кобозева, 2004], которые используются, например, при синтактико-семантическом анализе, для построения семантической структуры предложения и др.

При использовании в автоматическом анализе текстов ресурсов в виде семантических сетей, например, WordNet [Fellbaum, 1998] такие семантические классы образуются естественным образом – за счет понятий (синсетов) на верху иерархии семантической сети (суперпонятий, *supersenses*). В WordNet такой список суперпонятий включает 26 синсетов существительных, например, ANIMAL (животное), PERSON (человек), FEELING (чувство) и 15 синсетов глаголов – COMMUNICATION (общение), MOTION (движение), COGNITION (познание).

В работе [Flekova and Gurevich, 2016] авторы создают векторные представления для суперпонятий WordNet. Для этого они используют корпус страниц английской Википедии, автоматически размеченный по значениям ресурса BabelNet (включающего как составную часть WordNet) [Navigli and Ponzetto, 2012] – точность автоматической разметки оценивается как 77.8% [Scozzafava et al., 2015]. Размеченный корпус используется следующим образом: неоднозначные слова получают дополнительную пометку суперпонятия, к которому они относятся. На основе полученного корпуса с разметкой суперпонятий авторы обучают векторные представления слов, слов с разрешенной неоднозначностью относительно суперпонятия и самих суперпонятий в одном векторном пространстве *word2vec* [Mikolov et al., 2013b]. В результате эмбединга для слов, обученных совместно с суперпонятиями, позволили добиться более высокого качества предсказания семантической близости слов на четырех датасетах из пяти. Кроме того, эмбединги с суперпонятиями позволили улучшить результаты в задачах анализа тональности, классификации субъективных предложений и идентификации метафор.

В работе [Aloui et al., 2020] исследуется порождение контекстуализированных эмбедингов для суперпонятий WordNet на основе нейросетевой модели BiLSTM. Авторы используют шесть обобщенных суперпонятий для представления существительных: Одушевленная сущность, Природный объект, Артефакт (созданный объект), Динамическая ситуация (например, действие), Статическая ситуация (например, состояние). Для каждого суперпонятия авторы отбирают по 200 соответствующих, а также случайным образом дополнительно отбираются негативные примеры, не относящиеся к данному суперпонятию. Полученные списки примеров используются для разметки корпуса, в котором слова из списков заменяются на позитивные или негативные теги каждого суперпонятия, получается псевдоаннотированный корпус.

На полученном корпусе обучается классификатор, который, получая слово и контекст, предсказывает вероятность отнесения слова к конкретному суперпонятию. Для классификации используется нейронная модель BiLSTM. В результате каждое слово получает некоторую величину в век-

торе размерности 6 суперпонятий. Таким образом, для каждого слова получается интерпретируемый вектор. Подход тестируется в задаче предсказания суперпонятия для слова на размеченном значениями французском корпусе и сравнивается с нейросетевой языковой моделью FlauBERT [Le et al., 2020], французским вариантом модели BERT [Devlin et al., 2019]. Показано, что на небольших объемах обучающих данных предложенный подход сравним с FlauBERT. Кроме того, метод намного быстрее и менее затратен по ресурсам, чем FlauBERT.

Как видно, в описанных работах суперпонятия (абстрактные семантические категории) заранее фиксированы. В нашей работе мы исследуем подход к гибкому (адаптивному) выделению суперпонятий для задачи или предметной области на основе соответствующего текстового корпуса и структуры соответствующих векторных представлений.

## 2. Идея адаптивных суперпонятий

Суперпонятия – это синсеты RuWordNet (или других ресурсов в форме семантических сетей), которые обладают следующими свойствами:

- суперпонятие должно иметь несколько гипонимов (видовых понятий), т.е. действительно являться обобщающим понятием для более конкретных понятий, представлять собой узел близких слов по тезаурусу,

- суперпонятие как узел близких по смыслу слов должно быть подтверждено и на основе некоторого представительного корпуса – а именно, слова из суперпонятий и связанных с ними понятий должны образовывать клику – граф, в котором все вершины связаны между собой. Связи между словами в клике считаются на основе косинусной меры близости эмбедингов по заданному корпусу, превышающей заданный порог. Например, синсет *город*, связанный в RuWordNet с большим количеством конкретных городов, соответствует клике из более, чем 200 названий-городов, чьи векторные представления близки друг к другу с косинусной мерой выше 0.4, посчитанной на основе корпуса *Agancum*.

Таким образом, суперпонятия – это семантические категории, выделенные людьми в семантических ресурсах, дополнительно подтвержденные высокой семантической близостью векторов, связанных с ними слов, что дает возможность адаптации набора суперпонятий на конкретную задачу или конкретную предметную область, т.е. предполагается, что для разных задач и связанных с ними корпусов данных набор суперпонятий должен различаться, отражая наиболее существенные категории в данной предметной области.

### 3. Алгоритм порождения суперпонятий

Алгоритм формирования суперпонятий включает следующие шаги:

- 1) Выделение из RuWordNet синсетов, которые имеют по крайней мере несколько видовых отношений (гипонимов) с другими синсетами – это кандидаты в суперпонятия;
- 2) Порождение множества слов (сематического хаба), соответствующих суперпонятию-кандидату – это все слова и выражения из синсета понятия в RuWordNet, а также слова и выражения из синсетов, непосредственно связанных с этим понятием отношениями гипоним-гипероним (родовидовыми отношениями);
- 3) Создание клики понятий для суперпонятия на основе косинусной близости эмбедингов синсетов хаба и заданного порога. Клики высчитываются с помощью алгоритма из пакета NetworkX<sup>1</sup>. Вес связи между двумя синсетами определяется как средний вес связи слов из разных синсетов пары. Веса связей слова – это значение косинусной близости, посчитанной по некоторому корпусу дистрибутивной моделью (word2vec, glove). Усредненный вес должен превышать заданный порог. Чтобы был отобран как кандидат в суперпонятия, должна возникнуть минимальная клика – 3 понятия.
- 4) Для отобранного суперпонятия нужно создать представительное векторное представление, поскольку в семантическом хабе, связанном с суперпонятием (синонимы, гипонимы, гиперонимы) могут встречаться многозначные и/или редкие слова. Для этого из слов хаба создается сеть дистрибутивных отношений с весами косинусной близости, на сети вычисляется PageRank [Brin and Page, 1998] каждого слова, отбираются пять наиболее весомых слова, и их усредненное векторное представление рассматривается как векторное представление суперпонятия. Примеры отобранных суперпонятий на основе векторной модели корпуса Araneum [Kutuzov and, Kunilovskaya, 2017] представлены в табл. 1.
- 5) Следующим этапом является отбор суперпонятий из всех выделенных кандидатов. Отбор начинается с верхнего уровня иерархии семантической сети RuWordNet, поскольку мы заинтересованы в наиболее обобщающих, абстрактных понятиях. Если очередное суперпонятие похоже по векторной модели на уже отобранное и являющееся его гиперонимом, то это понятие отвергается, поскольку слишком много похожих суперпонятий приводят к запутанной системе семантических категорий.

---

<sup>1</sup> [https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.cliques.find\\_cliques.html#networkx.algorithms.cliques.find\\_cliques](https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.cliques.find_cliques.html#networkx.algorithms.cliques.find_cliques).

**Примеры выделенных суперпонятий и их наиболее  
представительных слов**

Идентификатор синсета RuWordNet	Название синсета RuWordNet	5 наиболее представительных слов, соответствующих суперпонятию
106509-N	Составная часть	<i>кусочек, полосочка, низ, полоска, бок.</i>
106501-N	Занятие, деятельность	<i>забава, занятие, развлечение, увлечение, отрасль</i>
106768-N	Свойство, характеристика	<i>абстрактность, безыскусность, стильность, отвлеченность, обособленность</i>
106646-N	Отношение между сущностями	<i>различие, соотношенность, взаимосвязь, несходство, несхожесть</i>
106505-N	Продукт труда	<i>подлинник, продукция, произведение, товар, издание</i>

Упрощенный фрагмент лога отбора суперпонятий выглядит следующим образом:

```
{'106509-N', 'name': 'составная часть', 'status': 'added'}
{'106646-N', 'name': 'отношение между сущностями', 'status': 'added'}
{'106846-N', 'name': 'вариант, разновидность', 'status': 'added'}
{'132964-N', 'name': 'новшество', 'status': 'skipped', 'condition':
'clique_size < 3'}
..
{'112766-N', 'name': 'задняя часть', 'status': 'skipped', 'condition': 'close
to (106509-N:0.59)'}
{'110735-N', 'name': 'частица (небольшая часть)', 'status': 'skipped',
'condition': 'close to (106509-N:0.50)'}
```

Из лога видно, что часть суперпонятий не одобряется (skipped) из-за отсутствия соответствующей клики (*'clique\_size < 3'*), а часть суперпонятий отвергается из-за слишком высокого сходства с уже отобранными суперпонятиями (*'close to'*).

Несмотря на сделанный отбор суперпонятий, выделяются еще близкие по смыслу суперпонятия, которые обычно представляют собой подвиды более широкого понятия (понятия-сестры), например, понятия *гуманитарные науки* и *общественные науки*. В таком случае естественно обобщать понятия-сестры к более широкому понятию. Приведем примеры обобщений похожих по векторной модели понятий-сестер к родовому понятию:



- *гуманитарные науки, общественные науки → наука*
- *научное учреждение, высшее учебное заведение → организация, выполняющая исследования и разработки*
- *иностранное государство, европейская страна, развитое государство, республика (государство), социалистическое государство → государство*
- *цветковое растение, травянистое растение, кустарник, вечнозеленое растение → растение*
- *воинское звание, дворянский титул → звание*

В зависимости от корпуса, на котором рассчитывается векторная модель, описанная процедура дает 100-300 суперпонятий, объединяющих разное количество элементов. Суперпонятий (семантических классов) не должно быть много. Поэтому на последнем шаге суперпонятия пополняются всеми остальными словами из корпуса (на текущем этапе только существительными из RuWordNet, присутствующими в корпусе), косинусное сходство которых с суперпонятиями выше некоторого порога, и отбирается заданное количество суперпонятий с максимальным количеством элементов.

Таким образом, в результате описанной процедуры отобраны суперпонятия (семантические категории) из ресурса RuWordNet, которым одновременно соответствует некоторое множество слов, близких между собой по корпусной дистрибутивной модели (клика). Каждое суперпонятие имеет свое векторное представление по исходной дистрибутивной модели. Все остальные слова из текстового корпуса можно представить в виде вектора схожести на эти суперпонятия, т.е. если отобрано 100 суперпонятий, то для каждого слова будет получен вектор размерности 100 интегрируемых измерений.

#### 4. Эксперименты

В эксперименте по описанному алгоритму использовался тезаурус RuWordNet и векторные модели, обученные на корпусе из 800 тысяч новостных статей. На основе близости слов к векторам суперпонятий каждому слову из векторной модели был сопоставлен вектор из  $K$  компонент по векторным близостям слова к суперпонятиям.

В данном эксперименте применялись следующие параметры:

- 1) Параметры алгоритма построения списка суперпонятий:
  - а. Минимальное количество родовидовых отношений у понятия для формирования суперпонятия (=5);

- b. Минимальное значение близости между понятиями для формирования ребра в процессе построения клики суперпонятия (=0.4);
  - c. Минимальное значение близости между словами для формирования ребра между словами в суперпонятии для построения графа и последующего расчета pagerank на нем (=0.4);
  - d. Минимальный размер клики из понятий в суперпонятии (=3);
  - e. Минимальный размер суперпонятия в словах (=10);
  - f. Количество слов для расчета вектора суперпонятия на основе PageRank (=5);
  - g. Минимальное значение близости между добавляемым суперпонятием и его уже добавленными гиперонимами при котором суперпонятие будет добавлено в итоговый список (=0.3);
- 2) Параметры алгоритма обобщения:
- a. Минимальное значение близости между парой ко-гипонимов для обобщения на гипероним (=0.4);
  - b. Минимальная глубина суперпонятия, которое может служить обобщением (=3);
  - c. Количество слов для расчета вектора суперпонятия (=5).

Приведем примеры полученных суперпонятий RuWordNet на данных параметрах: *составная часть; продукт труда; занятие, деятельность; пребывать в состоянии; изображение (результат); группа, объединенная по общему признаку; субъект деятельности; вещество; научный работник и др* (см. Приложение). Как видно, получилось выделить обобщающие семантические категории. Также можно заметить, что выделяются и более узкие категории, например, среди понятия *субъект деятельности* отдельно выделяется категория *научный работник*, которая имеет специфические контексты. В этом и состоит предполагаемая адаптивность метода выделения суперпонятий, который может выделить и существенные, важные, специфичные категории не обязательно самого верхнего уровня семантической сети.

В качестве эксперимента проверялась возможность отображать разные модели в одно интерпретируемое векторное пространство. Для этого на основе вышеупомянутого корпуса новостей были обучены две векторные модели word2vec с одинаковыми параметрами (skip-gram, размер окна 3, количество эпох 10).

Первой частью эксперимента была проверка, что две векторные модели, обученные даже на идентичном корпусе и с одинаковыми параметрами, будут иметь разные векторные представления. Для этого, для каждого слова из коллекции считалась косинусная близость между векторами из

двух разных моделей. Гистограмма близостей слов представлена на рис. 1. Средняя близость оказалась равна 0.6. Это означает, что обычные векторные модели, даже обученные на одном корпусе, создают несопоставимые векторные представления, в которых представления даже одного и того же слова могут быть не похожи друг на друга.

*Рис. 1. Гистограмма близостей слов к самим себе в разных моделях.  
По  $x$  – мера близости, по  $y$  – количество слов*

Для проверки сохранения семантической близости на основе векторов суперпонятий мы исследовали, насколько такое представление может выявлять сходство между семантически близкими словами из независимо обученных векторных моделей. Для этого предварительно был составлен список из  $\sim 100$  пар слов, на основе ранее обученной модели word2vec в той же постановке, которые имели высокое сходство. С этой целью для каждого слова из коллекции был получен список наиболее близких слов, эти списки были упорядочены по наивысшему сходству среди полученных близких слов. Для первых 100 слов в списке были выбраны их наиболее близкие слова.

Используя полученный список, при обучении векторных моделей, в первом случае из корпуса удалялись все «правые» слова пар, во втором случае все «левые» слова пар. Таким образом, итоговые словари векторных моделей отличались в соответствии с полученным списком пар. Например, для пары (**беларусь**, **белоруссия**) при обучении первой векторной модели, были удалены все случаи использования слова **белоруссия**, а при обучении второй модели, все случаи использования **беларусь**. Таким образом, эти модели не имеют в своих словарях соответствующих слов.

Далее, для каждой из двух векторных моделей, в соответствии с изложенным алгоритмом построения, были получены векторные представления на основе суперпонятий. Для того, чтобы полученные векторные модели были полностью сопоставимы, на нескольких этапах алгоритма было выполнено выравнивание:

1. Вес максимальной клики для каждого суперпонятия был получен усреднением соответствующих весов в двух моделях.
2. Итоговый список суперпонятий формировался как пересечение списков суперпонятий, полученных в двух моделях, а позиция суперпонятий формировалась как средняя позиция.

Таким образом, полученные списки суперпонятий получились идентичными и содержали в себе 78 суперпонятий. Так как итоговое количество суперпонятий из-за процедуры пересечения списков получилось меньше 100, то операция по усечению множества не проводилась. В более простом случае, когда работа идет только с одной моделью, обычно количество суперпонятий получается более 100.

В итоге каждое слово из модели было отображено в векторную модель на основе суперпонятий, и, если слово встречалось в обеих моделях, его представление усреднялось. В полученной векторной модели словарь является объединением словарей двух моделей.

Для проверки гипотезы, что с помощью разработанного подхода возможно отображать разные модели в одно векторное пространство был проведен анализ того, насколько близко к друг другу располагаются слова из выделенных ранее пар, по отношению к другим словам. Для этого, в полученной векторной модели для каждого слова из пары производилось вычисление списка наиболее близких слов, среди которых происходил поиск второго слова из пары и его позиции в списке. Например, 5 наиболее близких слов к слову **беларусь**: белоруссия 0.991, казахстан 0.988, судан 0.984, украина 0.984, россия 0.984, а к слову **белоруссия**: беларусь 0.991, казахстан 0.984, латвия 0.983, венгрия 0.983, молдавия 0.983. Таким образом, средний ранг для пары (**беларусь**, **белоруссия**) равен 0 (отсчет идет с 0).

Результаты показали, что для 60% пар, искомое слово оказывалось в топ-5. Итоговое распределение позиций можно видеть на рис. 2.

*Рис. 2: Гистограмма распределения рангов для исследуемых пар слов.  
По  $x$  – ранг, по  $y$  – количество пар*

Таким образом, на корпусе были независимо обучены две модели, в каждой из которых не содержалось некоторое множество слов, близких друг к другу, но в представлении на основе суперпонятий эти слова оказались близки.

## Заключение

В данной работе представлен подход по созданию интерпретируемых векторных представлений слов, в которых каждая компонента вектора соответствует некоторой интерпретируемой семантической категории. Для получения таких категорий используется лексико-семантический ресурс в виде семантической сети русского языка RuWordNet, а также представительный корпус русскоязычных текстов для порождения векторных представлений.

Полученные интерпретируемые векторные представления были протестированы на наличие возможности отображать разные модели в одно векторное пространство таким образом, чтобы близкие по смыслу слова, которые присутствуют в одной модели, но отсутствуют в другой, оказывались близко друг к другу в итоговой векторной модели.

## Список литературы

- [Апресян, 1974] Апресян Ю.Д. Лексическая семантика: Синонимические средства языка. – М., 1974.
- [Кобозева, 2004] Кобозева И. М. Лингвистическая семантика. – М., УРСС, 2004.
- [Agirre et al., 2009] Agirre E., Alfonseca, E., Hall, K., Kravalová, J., Pasca, M., & Soroa, A. A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches // Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2009.
- [Aloui et al., 2020] Aloui, C., Ramisch, C., Nasr, A., and Barque, L. Slice: Supersense-based lightweight interpretable contextual embeddings // In Proceedings of the 28th International Conference on Computational Linguistic, 2020.
- [Artetxe et al., 2009] Artetxe M., Labaka G. and Agirre E. Learning principled bilingual mappings of word embeddings while preserving monolingual invariance // Proceedings of the 2016 conference on empirical methods in natural language processing, 2016..
- [Brin and Page, 1998] Brin S. and, Page L. The anatomy of a large-scale hypertextual web search engine // Computer networks and ISDN systems. – 1998. – Vol. 30, No. 1-7.
- [Devlin et al., 2019] Devlin J., Chang M. W., Lee K., and Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. – Vol. 1 (Long and Short Papers). – 2019.

- [**Fellbaum, 1998**] Fellbaum C. WordNet: An Electronic Lexical Database. – Cambridge, MA., MIT Press, 1998.
- [**Flekova and Gurevich, 2016**] Flekova L., Gurevych I. Supersense embeddings: A unified model for supersense interpretation, prediction, and utilization // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). – 2016.
- [**Kutuzov and, Kunilovskaya, 2017**] Kutuzov A., Kunilovskaya M. Size vs. structure in training corpora for word embedding models: Araneum russicum maximum and russian national corpus //International Conference on Analysis of Images, Social Networks and Texts. – Springer, Cham, 2017.
- [**Le et al., 2020**] Le H., Vial L., Frej J., Segonne V., Coavoux M., Lecouteux B., ... and Schwab D. FlauBERT: Unsupervised Language Model Pre-training for French // In Proceedings of the 12th Language Resources and Evaluation Conference. – 2020.
- [**Leviant and Reichart, 2015**] Leviant I., Reichart R. Separated by an un-common language: Towards judgment language informed vector space modeling //arXiv preprint arXiv:1508.00106,2015.
- [**Loukachevitch et al., 2016**] Loukachevitch N.V. Lashevich G., Gerasimova A.A., Ivanov V.V., and Dobrov, B.V. Creating Russian wordnet by conversion // Computational Linguistics and Intellectual Technologies: papers from the Annual conference “Dialogue, 2016.
- [**Mikolov et al., 2013a**] Mikolov T., Le Q.V., Sutskever I. Exploiting similarities among languages for machine translation // arXiv preprint arXiv:1309.4168,2013.
- [**Mikolov et al., 2013b**] Mikolov T., Sutskever I., Chen K., Corrado G., and Dean J. Distributed representations of words and phrases and their compositionality // In Advances in Neural Information Processing Systems. – 2013. – Vol. 26.
- [**Navigli and Ponzetto, 2012**] Navigli R., Ponzetto S. P. BabelNet: Building a very large multilingual semantic network // Proceedings of the 48th annual meeting of the association for computational linguistics. – 2010.
- [**Panchenko et al., 2016**] Panchenko A., Ustalov D., Arefyev N., Paperno D., Konstantinova N., Loukachevitch N., and Biemann C. Human and machine judgements for Russian semantic relatedness // In International conference on analysis of images, social networks and texts. – 2016.
- [**Radovanovic et al., 2010.**] Radovanovic M., Nanopoulos A., Ivanovic M. Hubs in space: Popular nearest ' neighbors in high-dimensional data // The Journal of Machine Learning Research. – 2010.
- [**Ruder et al., 2019**]. Ruder S., Vulić I., Søgaard A. A survey of cross-lingual word embedding models // Journal of Artificial Intelligence Research. – 2019. – Vol. 65.
- [**Scozzafava et al., 2015**] Scozzafava F., Raganato A., Moro A., & Navigli R. Automatic identification and disambiguation of concepts and named entities in the multilingual Wikipedia // In Congress of the Italian Association for Artificial Intelligence. – 2015.
- [**Yamane et al. 2016**] Yamane J., Takatani T., Yamada H. и др. Distributional hypernym generation by jointly learning clusters and projections // Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. – 2016.

## Приложение

В приложении приводится список суперпонятий для модели, которая была предложена в описанном эксперименте:

- Суперпонятие 0: составная часть: полосочка, кусочек, мякоть, бочок, компонент
- Суперпонятие 1: продукт труда: произведение, продукция, товар, продукт, изделие
- Суперпонятие 2: занятие, деятельность: занятие, отрасль, хобби, профессия, увлечение
- Суперпонятие 3: пребывать в состоянии: томление, жизнелюбие, преклонение, оттаивание, голодание
- Суперпонятие 4: изображение (результат): фотография, рисунок, изображение, гравюра, снимка
- Суперпонятие 5: группа, объединенная по общему признаку: тройка, тридцатка, десяток, сотня, двадцатка
- Суперпонятие 6: субъект деятельности: должник, оппонент, покупатель, благотворитель, организатор
- Суперпонятие 7: вещество: отдушка, консервант, порошок, ингредиент, вещество
- Суперпонятие 8: измениться, изменение: улучшение, увеличение, ослабление, замедление, пропитка
- Суперпонятие 9: место в пространстве: стиб, выпуклость, пятнышко, поверхность, стремнина
- Суперпонятие 10: природное явление: погода, осадка, облако, ветер, циклон
- Суперпонятие 11: истратить, израсходовать: расход, экономия, потребление, перерасход, затрата
- Суперпонятие 12: компьютерная программа: утилитка, приложение, подпрограмма, программа, апдейт
- Суперпонятие 13: биологическая сущность: клетка, ткань, митохондрия, организм, хромосома
- Суперпонятие 14: движение, перемещение: катание, передвижение, наклон, путешествие, скачок
- Суперпонятие 15: состояние, внутренние обстоятельства: упорядоченность, сухость, влага, мокрота, ситуация
- Суперпонятие 16: население: горожанин, латиноамериканка, европейка, уралец, свердловчанин
- Суперпонятие 17: физиологический процесс: климакс, регенерация, секреция, сердцебиение, менопауза
- Суперпонятие 18: физическое свойство: проницаемость, густота, проводимость, ломкость, свойство
- Суперпонятие 19: единица измерения информации: мбайт, гигабайт, мбитый, кбит, мегабайт
- Суперпонятие 20: единица объема: галлон, декалитр, поллитра, баррель, литр
- Суперпонятие 21: изменить, сделать иным: улучшение, осветление, увеличение, уменьшение, изменение
- Суперпонятие 22: единица массы: килограмм, грамм, кило, тонна, унция
- Суперпонятие 23: способность: тактичность, музыкальность, умение, гибкость, расположенность
- Суперпонятие 24: материал для изготовления: поролон, латекс, пластик, обмотка, резина
- Суперпонятие 25: единица длины: микрометр, метр, сантиметр, нанометр, миллиметр
- Суперпонятие 26: сооружение, постройка: сооружение, здание, постройка, шалаш, строение
- Суперпонятие 27: предмет, вещь: мотка, брусочек, вещница, штурковина, украшение
- Суперпонятие 28: возраст: отрочество, детство, возраст, свойство, старость
- Суперпонятие 29: форма, наружный вид: округлость, выпуклость, комплектация, покрой, фасон
- Суперпонятие 30: бог: бог, господь, христос, богочеловек, иисус
- Суперпонятие 31: метод лечения: гемодиализ, терапия, диализ, рефлексотерапия, анестезия
- Суперпонятие 32: высказывание (то, что высказано): высказывание, рассуждение, реплика, шуточка, изречение
- Суперпонятие 33: чувство, эмоция: чувство, переживание, гордыня, разочарование, радость
- Суперпонятие 34: черта характера: сердечность, интеллигентность, уступчивость, уравновешенность, непоколебимость
- Суперпонятие 35: географическая область: приангарий, зауралье, урал, приамурья, прибайкалье
- Суперпонятие 36: создать (сделать существующим): протирание, кройка, отливка, ткачество, прокалывание
- Суперпонятие 37: небесное тело: комета, астероид, планета, солнце, луна
- Суперпонятие 38: административно-территориальная единица: англия, шотландия, волость, уезд, уэльс
- Суперпонятие 39: произнести, выговорить, проговорить: разглаживание, причитание, произнесение, трескотня, акцентирование
- Суперпонятие 40: мнение: трагичность, трагизм, предположение, суждение, истинность
- Суперпонятие 41: исследование, изучение: изучение, исследование, анализ, изыскание, слежение
- Суперпонятие 42: промежуток времени: месяц, час, полгода, век, полчаса
- Суперпонятие 43: вооруженное формирование: спецназ, отряд, подразделение, войско, спецотряд
- Суперпонятие 44: важный, значительный: важность, всемогущество, могущество, величие, приоритетность
- Суперпонятие 45: положение, тезис: догмат, постулат, суждение, мысль, принцип
- Суперпонятие 46: переместить: опускание, переброс, тряска, размахивание, сваливание
- Суперпонятие 47: государство: индия, джибути, египет, кения, малайзия
- Суперпонятие 48: созвездие зодиака: водолей, овен, телец, скорпион, стрелец

Суперпонятие 49: высшее учебное заведение: вуз, сельхозинститут, университет, госуниверситет, факультет  
Суперпонятие 50: воинское звание: лейтенант, сержант, подполковник, майор, ефрейтор  
Суперпонятие 51: стиль в искусстве: модернизм, романтизм, минимализм, реализм, классицизм  
Суперпонятие 52: предприятие общественного питания: ресторан, кафе, ресторанич, бистро, кафешка  
Суперпонятие 53: массовая акция: шествие, митинг, пикет, акция, марш  
Суперпонятие 54: физические упражнения: упражнение, трусца, подтягивание, гимнастика, разминка  
Суперпонятие 55: скульптурное произведение: изваяние, монумент, скульптура, памятник, статуя  
Суперпонятие 56: полуостров: камчатка, таймыр, полуостров, аляска, флорида  
Суперпонятие 57: изобразить (воспроизвести): вышивание, резьба, тиснение, гравировка, вышивка  
Суперпонятие 58: издать звук: урчание, завывание, шипение, дрожание, шуршание  
Суперпонятие 59: достижения в деятельности: успех, достижение, вклад, итог, лепта  
Суперпонятие 60: кормить, давать еду: корм, прикорм, подкормка, откорм, прикормка  
Суперпонятие 61: план деятельности: программа, госпрограмма, подпрограмма, перечень, список  
Суперпонятие 62: марка автомобиля: газель, мазд, ситройня, жигули, фольксваген  
Суперпонятие 63: приготовление пищи: выпекание, запекание, выпечка, копчение, варка  
Суперпонятие 64: военнослужащий: военнослужащий, танкист, кавалерист, офицер, десантник  
Суперпонятие 65: водные средства транспорта: судно, корабль, теплоход, катамаран, буксир  
Суперпонятие 66: экономический показатель: рентабельность, пассажирооборот, грузооборот, себестоимость, показатель  
Суперпонятие 67: деньги: деньга, сумма, расход, оплата, задолженность  
Суперпонятие 68: драка: драка, потасовка, стычка, конфликт, поножовщина  
Суперпонятие 69: перелетная птица: куропатка, чирок, скворец, аист, перепел  
Суперпонятие 70: город иркутской области: слюдянка, шелехов, иркутск, нижнеудинск, черемхово  
Суперпонятие 71: научный работник: естествоиспытатель, историк, профессор, этнолог, филолог  
Суперпонятие 72: курортный город: кисловодск, пятиягорск, сочи, евпатория, железнодорожск  
Суперпонятие 73: охотничья собака: собака, собачонка, собачка, спаниель, терьер  
Суперпонятие 74: заместитель должностного лица: заместитель, замглавы, зам, замминистра, замначальник  
Суперпонятие 75: награда: вознаграждение, награда, медаль, приз, премия  
Суперпонятие 76: записать на специальный носитель: фиксация, видеозапись, запись, аудиозапись, видеофиксация  
Суперпонятие 77: рыба: толстолобик, камбала, рыба, скумбрия, корюшка



---

СЕКЦИЯ 2

**ИНТЕЛЛЕКТУАЛЬНЫЙ  
АНАЛИЗ ДАННЫХ  
И ОБЪЯСНИМЫЙ  
ИСКУССТВЕННЫЙ  
ИНТЕЛЛЕКТ**

УДК 004.415.2

**ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ  
ОБЪЯСНИТЕЛЬНОГО ИСКУССТВЕННОГО  
ИНТЕЛЛЕКТА В ЗАДАЧАХ  
ПЕРСОНАЛИЗИРОВАННОЙ МЕДИЦИНЫ\***

А.Н. Аверкин (*averkin2003@inbox.ru*)

С.А. Ярушев (*Yarushev.sa@rea.ru*)

Российский экономический университет  
им. Г.В. Плеханова, Москва

Исследования в области объяснительного искусственного интеллекта активно ведутся во всем мире, в том числе и в России. Разработка объяснительного аппарата крайне необходима, так как использование искусственного интеллекта (ИИ) в критически значимых областях, таких как здравоохранение, накладывает дополнительные требования к надежности систем данного класса. Необходимо повысить уровень доверия к ИИ, путем повышения уровня прозрачности решений, принимаемых ИИ. В данной работе проведен обзор наиболее перспективных областей здравоохранения, в

---

\* Исследование выполнено за счет гранта Российского научного фонда № 22-71-10112, <https://rscf.ru/project/22-71-10112/>.

которых могут применяться системы, основанные на искусственном интеллекте. Так же, были рассмотрены различные нейро-нечеткие модели искусственных нейронных сетей, потенциально перспективные для разработки на их основе объяснительных систем искусственного интеллекта.

**Ключевые слова:** искусственный интеллект, ХАИ, здравоохранение, нейронные сети, объяснительный искусственный интеллект, нейро-нечеткие сети.

## Введение

В мире в основном исследования в области разработки объяснительного искусственного интеллекта связаны с системами объяснений третьего поколения и начавшейся в 2018 г. программой DARPA (Defense Advanced Research Projects Agency) [Gunning et al., 2019]. Программа DARPA по созданию систем объяснительного ИИ стремится создать такие системы ИИ, чьи модели обучения и решения могут быть понятны и должным образом проверены конечными пользователями. Достижение этой цели требует методов построения более объяснимых моделей, разработки эффективных объяснимых интерфейсов и понимания психологических требований для эффективного объяснения. Объяснительный ИИ нужен для того, чтобы пользователи понимали, должным образом доверяли и эффективно управляли своими умными партнерами. DARPA рассматривает ХАИ (Explainable Artificial Intelligence) как системы ИИ, которые могут объяснить свое решение человеку-пользователю, охарактеризовать свои сильные и слабые стороны, и то, как они будут вести себя в будущем. Целью DARPA является создание более понятных для человека систем ИИ с помощью эффективных объяснений. Команды разработчиков ХАИ решают первые две проблемы путем создания и развития технологий объяснимого машинного обучения, разрабатывая принципы, стратегии и методы взаимодействия человека и компьютера для получения эффективных объяснений. Еще одна команда разработчиков ХАИ решает третью задачу путем объединения, расширения и применения психологических теорий объяснения, которые команды разработчиков будут использовать для тестирования своих систем. Команды разработчиков оценивают, насколько хорошо объяснения ХАИ-систем улучшают работу пользователей, их доверие и производительность.

В России также уделяется большое внимание направлению объяснимого ИИ. Так, Нижегородский государственный университет в 2020 г. стал победителем в конкурсе крупных научных проектов от Минобрнауки РФ с проектом «Надежный и логически прозрачный искусственный ин-

теллект: технология, верификация и применение при социально-значимых и инфекционных заболеваниях» [Горбань, 2020]. Главным результатом проекта должна стать разработка новых методов и технологий, позволяющих преодолеть два основных барьера систем машинного обучения и ИИ: проблему ошибок и проблему явного объяснения решений. Руководитель проекта профессор Александр Горбань так объяснил основную идею проекта: «Эти проблемы тесно связаны: без возможности логического прочтения ошибки ИИ будут оставаться необъяснимыми. Дообучение системы в рамках существующих методов может повредить имеющиеся навыки и, с другой стороны, может потребовать огромных ресурсов, что в серьезных задачах непрактично. К примеру, широко известная система когнитивных вычислений IBM Watson потерпела неудачу на рынке персонализированной медицины вследствие систематически совершаемых ошибок в диагностике и рекомендации лечения рака, найти и устранить источники которых не удалось». Далее, мы рассмотрим возможности применения нейро-нечетких сетей для разработки объяснительных архитектур искусственного интеллекта. В параграфе 2 проведен обзор наиболее перспективных областей применения систем искусственного интеллекта в сфере здравоохранения.

## **1. Объяснительный искусственный интеллект на основе нейро-нечетких сетей**

Наиболее интересным направлением в области разработки объяснительных моделей ИИ является извлечение правил с использованием нейронечетких моделей. Системы, основанные на нечетких правилах, разработанные с помощью нечеткой логики, стали полем активных исследований за последние несколько лет. Эти алгоритмы доказали свои сильные стороны в таких задачах, как управление сложными системами, создание нечетких элементов управления. Взаимоотношения между обоими подходами были тщательно изучены и показана их эквивалентность. Это позволяет сделать два важных вывода. Во-первых, можно применить то, что было обнаружено для одной из моделей, к другой. Во-вторых, мы можем перевести знания, встроенные в нейронную сеть, на более когнитивно-приемлемый язык - нечеткие правила. Другими словами, получаем семантическую интерпретацию нейронных сетей [Averkin et al., 2018].

Для того, чтобы получить семантическую интерпретацию черного ящика глубокого обучения, нейро-нечеткие сети могут быть использованы вместо последнего полносвязного слоя. Например, ANFIS (адаптивная нейронечеткая сеть) является многослойной сетью прямого распространения. Эта архитектура имеет пять слоев, таких как нечеткий слой, продук-

ционный слой, слой нормализации, слой дефаззификации и выходной слой. ANFIS сочетает преимущества нейросети и нечеткой логики. Далее приведем классификацию наиболее известных нейронечетких подходов.

Рассматривая архитектуры нейронечетких моделей, можно выделить три методики объединения искусственных нейронных сетей (ИНС) и нечетких моделей [Jin, 2011]:

- neuro-FIS (neuro-fuzzy inference systems), в которых ИНС используется как инструмент в нечетких моделях;
- нечеткие ИНС, в которых классические модели ИНС фаззифицированы;
- нейро-нечеткие гибридные системы, в которых нечеткие системы и ИНС объединены в гибридные системы [Борисов, 2017].

Исходя из данных методик, нейро-нечеткие модели можно разделить на три класса [Rajab, 2018].

Кооперативные нейро-нечеткие модели. В данном случае часть ИНС изначально используется для определения нечетких множеств и / или нечетких правил, где впоследствии выполняется только полученная нечеткая система. В процессе обучения определяются функции принадлежности, а также формируются нечеткие правила на основе обучающей выборки. Здесь основная задача нейронной сети заключается в подборе параметров нечеткой системы.

Параллельные нейро-нечеткие модели. Нейронная сеть в данном типе модели работает параллельно с нечеткой системой, предоставляя входные данные в нечеткую систему или изменяя выходные данные нечеткой системы. Нейронная сеть может являться также и постпроцессором выходных данных из нечеткой системы.

Гибридные нейро-нечеткие модели. Нечеткая сеть использует метод обучения, как это делает и ИНС, чтобы настроить свои параметры на основе обучающих данных. Среди представленных классов моделей наибольшей популярностью пользуются модели именно данного класса, доказательством тому служит их применение в широком спектре реальных задач [Shihabudheen, 2018].

Среди наиболее популярных гибридных моделей можно выделить следующие архитектуры.

Сеть управления нечетким адаптивным обучением (FALCON) [Lin, 1991], которая имеет пятислойную архитектуру. На одну выходную переменную приходится по два лингвистических узла. Первый узел работает с обучающей выборкой (паттерном обучения), второй является входным для всей системы. Первый скрытый слой размечает входную выборку в соответствии с функциями принадлежности. Второй слой задает правила и их параметры. Обучение происходит на основе гибридного алгоритма

без учителя для определения функции принадлежности, базы правил и использует алгоритм градиентного спуска для оптимизации и подбора итоговых параметров функции принадлежности.

Адаптивная нейронечеткая сеть вывода ANFIS – это хорошо известная нейронечеткая сеть, которая применялась во многих приложениях и исследовательских областях [Jang, 1993]. Более того, сравнение архитектур нейронечетких сетей показало, что ANFIS показывает минимальную ошибку в задаче прогнозирования. Основным недостатком модели ANFIS является то, что она предъявляет серьезные требования к вычислительной мощности [Fan, 2017].

Сеть обобщенного приближенного интеллектуального управления на основе рассуждений (GARIC) [Bherenji, 1992] представляет собой нейронечеткую систему, использующую два нейросетевых модуля, модуль выбора действия и модуль оценки состояния, который отвечает за оценку качества выбора действий предыдущим модулем. GARIC - пятислойная сеть прямого распространения.

Нейронный нечеткий регулятор (NEFCON) [Nauck, 1999] был разработан для реализации системы нечеткого вывода типа Мамдани. Связи определяются с помощью нечетких правил. Входной слой является фазификатором, а выходной решает задачу дефазификации. Обучается сеть на основе гибридного алгоритма обучения с подкреплением и алгоритма обратного распространения ошибки.

Сеть нечеткого вывода и нейронной сети в программном обеспечении нечеткого вывода (FINEST) [Tano, 1996] представляет собой систему настройки параметров. Производится настройка нечетких предикатов, функции импликации и комбинаторной функции.

Система для автоматического построения нейронной сети нечеткого вывода (SONFIN) [Juang, 1998] по своей сути аналогична NEFCON контроллеру, но вместо реализации нечеткого вывода типа Мамдани реализует вывод типа Такаги-Сугено. В данной сети входная выборка обрабатывается с помощью алгоритма выровненной кластеризации. При идентификации структуры части предварительного условия входное пространство разделяется гибким образом в соответствии с алгоритмом, основанным на выровненной кластеризации.

Динамически развивающаяся нечеткая нейронная сети (dmEFuNN) и (EFuNN) [Kasabov, 1999]. В EFuNN все узлы создаются в процессе обучения. Первый слой передает обучающие данные на второй, который вычисляет степень соответствия с заранее определенной функцией принадлежности. Третий слой содержит в себе наборы нечетких правил, являющихся прототипами входных-выходных данных, которые можно представить в качестве гиперсфер нечеткого входного и выходного пространств.

Четвертый слой рассчитывает степень, с которой выходная функция принадлежности разметила входные данные, а пятый слой производит дефазификацию и подсчитывает числовые значения выходной переменной. DmEfuNN представляет собой модифицированную версию EFuNN. Основная идея состоит в том, что для всех входных векторов динамически подбирается набор правил, значения активации которых используются для расчета динамических параметров выходной функции. В то время как EFuNN реализует нечеткие правила типа Мамдани, dmEfuNN применяет тип Такаги-Сугено.

## 2. Искусственный интеллект в здравоохранении

Отрасль здравоохранения много лет инвестирует в технологические решения, включая искусственный интеллект. Разработка гибридных моделей, объединяющих разные технологические решения может синергетически улучшить положительные результаты. Например, решение ИИ может стать экспоненциально более мощным и надежным, если оно объединит в себе достижения из различных областей науки, в том числе, если оно будет объединено с дополненной реальностью, виртуальной реальностью, более быстрыми вычислительными системами, робототехникой или Интернетом вещей.

Голосовые помощники. Голосовые помощники могут участвовать в двустороннем диалоге с пользователем посредством распознавания речи, обработки естественного языка, понимания естественного языка и генерации естественного языка. За многими из них стоит ИИ. Подобные системы могут включать текстовые диалоги, разговорный язык и многое другое. Некоторые голосовые помощники имитируют человеческий образ (например, образ медсестры или тренера) или иной образ (например, животное), чтобы добавить интерактивности и разнообразить беседу.

Мониторинг здоровья и прогнозирование рисков. ИИ может использовать необработанные данные с акселерометров, гироскопов, микрофонов, камер и других датчиков, включая смартфоны. Алгоритмы машинного обучения можно обучить распознавать закономерности на основе исходных данных, а затем классифицировать эти закономерности как индикаторы поведения и состояния здоровья человека. Подобные системы позволяют пациентам следить за своим здоровьем, выявлять различные симптомы и понимать их, а также обмениваться данными с медицинскими учреждениями.

В настоящее время популярность подобных умных устройств, интеллектуальных устройств и мобильных приложений для здоровья резко возросла. Всего за 4 года, с 2014 по 2018 год, доля взрослых в США, использующих умные устройства для контроля здоровья, увеличилась с 9 до 33

процентов. Использование мобильных приложений для здоровья увеличилось с 16 до 48 процентов. С 2013 года стартап-компании в сфере ИИ, специализирующиеся на здравоохранении и умных устройствах, собрали 4,3 миллиарда долларов на разработку, например, бюстгалтеров, предназначенных для прогнозирования риска рака груди, и умной одежды для определения сердечных заболеваний.

Своевременное индивидуальное вмешательство. Адаптивные помощники, основанные на искусственном интеллекте (АПИИ). Это обучающие системы, которые обеспечивают динамическое индивидуальное лечение пользователей в зависимости от времени и течения болезни. АПИИ принимает решения о том, когда и какие действия предпринимать, основываясь на реакции на предыдущие действия, а также на осознании текущего контекста, будь то внутренний (настроение, тревога, артериальное давление) или внешний (например, местоположение, активность). Помощь АПИИ предоставляется тогда, когда пользователи больше всего в ней нуждаются. Подобные системы также могут сообщить врачу об обнаружении некоего тревожного шаблона. Например, вмешательство АПИИ может определить, когда пользователь находится в опасной ситуации и незамедлительно предпринять необходимые меры.

Решения на основе ИИ для эффективного управления медицинскими учреждениями. Есть две основные области применения ИИ в клинической помощи: улучшение и оптимизация оказания медицинской помощи и повышение эффективности обработки информации, взаимодействия с пользователем и интеллектуальной поддержки в электронных медицинских картах. На протяжении многих лет в данных областях были достигнуты успехи, в основном за счет основанных на правилах экспертных систем, обычно ориентированных на конкретные предметные области или проблемы.

Информационные кампании и кампании по вопросам здоровья на основе ИИ. ИИ может помочь определить конкретные демографические или географические районы, где распространены заболевания или высокий риск их распространения. Исследователям удалось успешно применить архитектуру сверточной нейронной сети для количественной оценки связи между средой обитания и распространенностью ожирения. Они показали, что физические характеристики района могут быть связаны с вариациями распространенности ожирения в разных районах. Так же, на основе машинного обучения, с применением таких инструментов как биомаркеры, так и социомаркеры, была разработана система для прогнозирования и идентификации пациентов с астмой у детей, которым грозит повторное посещение больниц.

Улучшение здоровья населения посредством ведения хронических заболеваний. Уже существует спектр готовых к использованию подходов, на основе искусственного интеллекта для поддержки программ здоровья населения. Они используются для автоматического скрининга сетчатки, поддержки принятия клинических решений и инструментов самоконтроля пациентов. Несколько подобных решений получили одобрение и применяются в системах здравоохранения западных стран, например, сервис под названием Medtronic's Guardian Connect, который стал первой системой непрерывного мониторинга уровня глюкозы на базе искусственного интеллекта. Другим примером подобных подходов являются собранные на основе краудсорсинга реальные данные об использовании ингаляторов в сочетании с данными об окружающей среде привели к созданию модели политических рекомендаций, которую можно воспроизвести для решения многих проблем общественного здравоохранения путем одновременного принятия индивидуальных, клинических и политических решений.

### **Заключение**

В данной работе были рассмотрены методы извлечения правил из искусственных нейронных сетей с целью создания на их основе систем объяснительного искусственного интеллекта. Извлечение правил является одним из способов помочь понять нейронные сети. Извлечение правил позволяет выявить скрытые зависимости в нейронных сетях, в случае, когда правила применяются для объяснения работы отдельных нейронов. Таким образом, благодаря применению методов извлечения правил, становится возможным объяснить принятие того или иного решения нейронной сетью на основе полученных правил. Также, в работе был приведен обзор интеллектуальных систем, задействованных в области здравоохранения, и в основе которых лежат методы искусственного интеллекта. В работе сделана попытка выявить наиболее перспективные области здравоохранения, в которых целесообразно внедрять объяснительные системы искусственного интеллекта. В результате проведенных исследований были выявлены области применения объяснительных систем искусственного интеллекта, проанализированы наиболее интересные, на взгляд авторов, методы извлечения правил из искусственных нейронных сетей. На основе проведенных исследований будут вестись дальнейшие работы по разработке объяснительных систем искусственного интеллекта и апробации их в области здравоохранения.



## Список литературы

- [Averkin et al., 2018] Averkin A., Yarushev S. Hybrid Neural Networks and Time Series Forecasting // Artificial Intelligence. Communication in Computer and Information Sciences. – 2018. – Vol. 934. – P. 230-239.
- [Bherenji, 1992] Bherenji H.R., Khedkar P. Learning and Tuning Fuzzy Logic Controllers through Reinforcements // IEEE Transactions on Neural Networks. – 1992. – Vol. (3). – P. 724-740.
- [Fan, 2017] Fan L. Revisit Fuzzy Neural Network: Demystifying Batch Normalization and Relu with Generalized Hamming Network // Proceedings of the 31st International Conference on Neural Information Processing Systems. – Long Beach California USA, 2017. – P. 1920-1929.
- [Gunning et al., 2019] Gunning, D., Aha, D. DARPA's Explainable Artificial Intelligence (XAI) Program // AI Magazine. – 2019. – Vol. 40 (2). – P. 44-58.
- [Jang, 1993] Jang J.-S. R. ANFIS: Adaptive-Network-Based Fuzzy Inference System // IEEE Trans. Systems & Cybernetics. – 1993. – Vol. 23. – P. 665-685.
- [Jin, 2011] Jin X.-H. Model for Efficient Risk Allocation in Privately Financed Public Infrastructure Projects Using Neuro-Fuzzy Techniques // J. Constr. Eng. Manag. – 2011. – P. 1003-1014.
- [Juang, 1998] Juang Chia Feng, Lin Chin Teng. An Online Self Constructing Neural Fuzzy Inference Network and its Applications // IEEE Transactions on Fuzzy Systems. – 1998. – Vol. 6. No. 1. – P. 12-32.
- [Kasabov, 1999] Kasabov N., Qun Song Dynamic Evolving Fuzzy Neural Networks with 'm-out-of-n' Activation Nodes for On-line Adaptive Systems. Technical Report TR99/04, Department of information science. University of Otago. – Otago, 1999.
- [Lin, 1991] Lin C.T., Lee C.S.G. Neural Network based Fuzzy Logic Control and Decision System // IEEE Transactions on Comput. – 1991. – Vol. 40 (12). – P. 1320-1336.
- [Nauck, 1999] Nauck D., Kruse R. Neuro-Fuzzy Systems for Function Approximation // Fuzzy Sets and Systems. – 1999. – Vol. 101 (2). – P. 261-271.
- [Rajab, 2018] Rajab S., Sharma V. A Review on the Applications of Neuro-Fuzzy Systems in Business // Artif. Intell. Rev. – 2018. – Vol. 49. – P. 481-510.
- [Shihabudheen, 2018] Shihabudheen K.V., Pillai G.N. Recent Advances in Neuro-Fuzzy System: A Survey // Knowl.-Based Syst. – 2018. – Vol. 152. – P. 136-162.
- [Tano, 1996] Tano S., Oyama T., Arnould T. Deep combination of Fuzzy Inference and Neural Network in Fuzzy Inference // Fuzzy Sets and Systems, – 1996. – Vol. 82 (2). – P. 151-160.
- [Борисов, 2017] Борисов В.В., Федулов А.С., Зернов М.М. Основы гибридизации нечетких моделей. Серия «Основы нечеткой математики». Кн. 9. – М.: Горячая линия–Телеком, 2017. – 100 с.
- [Горбань, 2020] Горбань А.Н. Ошибки интеллекта, основанного на данных // Интеллектуальные системы в науке и технике. Искусственный интеллект в решении актуальных социальных и экономических проблем XXI в. // Сб. ст. по матер. Междунар. конф. «Интеллектуальные системы в науке и технике» и Шестой всероссийской научно-практической конф. «Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века». – Пермь, 2020. – С. 11-13.

## МИНИМИЗАЦИЯ ЭМПИРИЧЕСКОГО РИСКА КАК СРЕДСТВО ВЫБОРА ЧИСЛА ВКФ-ГИПОТЕЗ

Д.В. Виноградов (*krrguest@yandex.ru*)  
Федеральный исследовательский центр  
«Информатика и управление» РАН, Москва

В работе исследуется новый подход к оценке числа требуемых гипотез о причинах целевого свойства. Он следует классическому методу В.Н. Вапника и А.Я. Червоненкиса – минимизации числа ошибок классификации на обучающей выборке. Для простейшего случая булевой алгебры без учета контр-примеров получен асимптотический результат о быстром убывании ошибок предсказания при возрастании объема обучающей выборки.

**Ключевые слова:** цепь Маркова, индуктивное обобщение, эмпирический риск, предсказание по аналогии.

### Введение

ВКФ-метод алгебраического машинного обучения был предложен и исследован ранее (см., например, [Vinogradov, 2018]). Он использует спаривающую цепь Маркова (СЦМ) для индуктивного обобщения (вычисления сходств) обучающих примеров в гипотезы о причинах их целевого свойства. Семейство СЦМ обеспечивает останавливаемость процесса вычислений с вероятностью единица. Ранее исследовались другие варианты: например, монотонная цепь Маркова (МЦМ), которая в случае булевой алгебры совпадает с классическим ленивым случайным блужданием на соответствующем булевом гиперкубе. Для случая булевой алгебры была получена [Виноградов, 2018а] точная оценка скорости перемешивания МЦМ порядка  $O(\frac{1}{2} \cdot n \cdot \log n)$ . При этом предельное распределение будет равномерным. Для случая булевой алгебры у спаривающей цепи Маркова были доказаны теоремы о сильной концентрации длин траекторий около среднего значения, которое имеет порядок  $O(n \cdot \log n)$ . Так как как нижняя компонента спаривающей цепи Маркова движется по траектории монотонной цепи Маркова, то эти результаты в совокупности приводят к тому,

что результаты выдачи СЦМ можно считать приближенно равномерно распределенными. Поэтому мы в дальнейшем будем предполагать, что гипотезы имеют равномерное распределение на булевом гиперкубе.

С помощью спаривающей цепи Маркова осуществляется индуктивное обобщение – вероятностное порождение [Vinogradov, 2018] ВКФ-гипотез о причинах целевого свойства через сходства обучающих примеров. Для устранения переобучения (так называемых «фантомных» гипотез [Виноградов, 2018b]) применяется условие «запрета контр-примеров». Ключевым параметром алгоритма является число ВКФ-гипотез, которых достаточно для минимизации числа ошибочных предсказаний целевого свойства у тестовых примеров. Алгоритм предсказания реализует рассуждение по аналогии [Vinogradov, 2018].

В работе [Виноградов, 2022] была получена улучшенная нижняя оценка  $\frac{n \cdot \ln 2 - \ln \delta}{\varepsilon}$  на число ВКФ-гипотез, чтобы с заданной надежностью (вероятностью  $1 - \delta$ , близкой к единице) были правильно предсказаны все  $\varepsilon$ -важные тестовые примеры. Этот подход дуален методу В.Н. Вапника – А.Я. Червоненкиса [Вапник и др., 1974] вероятно приближенно-корректного обучения.

В.Н. Вапник и А.Я. Червоненкис предложили и прямой подход к выбору хорошего классификатора – метод минимизации эмпирического риска. Мы превратим этот метод в средство получения достаточного числа ВКФ-гипотез.

Основной результат настоящей работы – получение оценки доли ошибочных предсказаний для ВКФ-метода в особенно простом случае, когда обучающие и тестовые примеры представляют собой коатомы в булевой алгебре, а контр-примеры не учитываются. Так как по предположению гипотезы имеют равномерное распределение, можно свести нашу ситуацию к сериям испытаний Бернулли и применить прямой подсчет.

## 1. Вспомогательные определения и факты

### 1.1. Элементы теории решеток

Наименьшая **обучающая выборка** для  $n$ -мерной булевой алгебры – это тройка  $\langle O, F, I \rangle$ , где  $O = \{o_1, \dots, o_n\}$  – множество коатомов (обучающих примеров),  $F = \{f_1, \dots, f_n\}$  – множество бинарных атрибутов (кодирующих признаки), и  $I \subseteq O \times F$  – отношение, определяемое как  $o_i I f_j \Leftrightarrow i \neq j$ .

**Кандидат** для обучающей выборки  $\langle O, F, I \rangle$  – это пара  $\langle A, B \rangle$ , где  $A' = \{f_j \in F : \forall o_i \in A[o_i I f_j]\} = B$  и  $B' = \{o_i \in O : \forall f_j \in B[o_i I f_j]\} = A$ .

Первая компонента  $A \subseteq O$  кандидата  $\langle A, B \rangle$  называется **списком родителей**, а вторая  $B \subseteq F$  – (общим) **фрагментом** кандидата. Частичный **порядок** между кандидатами задается правилом:  $\langle A, B \rangle \leq \langle C, D \rangle \Leftrightarrow B \subseteq D$ .

Случай булевой алгебры вынуждает нас не использовать прямой ДСМ-подход [Финн и др., 2009] к извлечению знаний из обучающих выборок большого объема (Big Data), когда сначала порождаются все кандидаты, а затем некоторые из них отбрасываются с помощью проверок дополнительных фальсифицирующих условий (например, «запрет контр-примеров»). Обучающая выборка требует  $n^2$  бит, а для сохранения всех кандидатов нужно  $n \cdot 2^n$  бит памяти.

Взамен мы предлагаем порождать с помощью вероятностных алгоритмов случайное подмножество кандидатов. Каждый кандидат порождается одной траекторией случайного блуждания через решетку кандидатов. Базовые шаги называются операциями «замыкай-по-одному» в честь соответствующего алгоритма С.О. Кузнецова [Кузнецов, 1993].

Операция  $CbODown$  заменяет кандидат  $\langle A, B \rangle$  относительно объекта  $o \in O$  на кандидат  $CbODown(\langle A, B \rangle, o) = \langle (B \cap \{o\})', B \cap \{o\}' \rangle$ . Операция  $CbOUp$  заменяет кандидат  $\langle A, B \rangle$  относительно атрибута  $f \in F$  на кандидат  $CbOUp(\langle A, B \rangle, f) = \langle A \cap \{f\}', (A \cap \{f\}')' \rangle$ .

Основное свойство этих операций выражается следующей леммой.

**Лемма 1.** Для любой упорядоченной пары кандидатов  $\langle A, B \rangle \leq \langle C, D \rangle$  и любого объекта  $o \in O$  выполняется неравенство  $CbODown(\langle A, B \rangle, o) \leq CbODown(\langle C, D \rangle, o)$ . Для любой упорядоченной пары кандидатов  $\langle A, B \rangle \leq \langle C, D \rangle$  и любого атрибута  $f \in F$  верно неравенство  $CbOUp(\langle A, B \rangle, f) \leq CbOUp(\langle C, D \rangle, f)$ .

## 1.2. Вероятностные алгоритмы порождения кандидатов

Случайные блуждания могут осуществляться несколькими разными способами. Первоначально это были алгоритмы, основанные на немонотонной и монотонной цепях Маркова. Основная проблема с ними (в случае произвольной обучающей выборки) была в отсутствии хорошей (полиномиальной) оценки на время перемешивания  $T$  (длину траектории, чтобы достичь близкого к стационарному распределению), которая служит для них входным параметром.

Для случая булевой алгебры эти алгоритмы совпадают со стандартным и ленивым случайным блужданием по булевому гиперкубу, соответственно. Стационарное распределение для них – равномерное.

Простейший вариант монотонной цепи Маркова составляет основу Алгоритма 1.

**Data:** множество обучающих примеров; функции  $CbODown(\cdot)$  и  $CbOUp(\cdot)$  операций «замыкай-по-одному», число шагов  $T$ .

**Result:** кандидат  $\langle A, B \rangle$  в гипотезы

$O :=$  обучающие примеры,  $F :=$  признаки;  $I \subseteq O \times F$  – выборка;

$R := O \cup F$ ;  $Min := \langle O, O' \rangle$ ;

```

for ( $i := 0; i < T; i++$ ) do
    Выбираем случайный элемент  $r \in R$ ;
    if ( $r \in O$ ) then
         $Min := CbODown(Min, r)$ ;
    else
         $Min := CbOUp(Min, r)$ ;
end
return  $Min$ ;

```

### Алгоритм 1: Монотонная цепь Маркова

Чтобы избежать проблемы с оценкой времени перемешивания в общем случае, автор (см., например, [Vinogradov, 2018]) предложил использовать семейство спаривающих цепей Маркова. Простейший вариант составляет собой Алгоритм 2, его состояниями являются упорядоченные пары кандидатов. Корректность алгоритма обосновывается леммой 1.

**Data:** множество обучающих примеров; функции  $CbODown(,)$  и  $CbOUp(,)$  операций «замыкай-по-одному»

**Result:** кандидат  $\langle A, B \rangle$  в гипотезы

$O :=$  обучающие примеры,  $F :=$  признаки;  $I \subseteq O \times F$  – выборка;

$R := O \cup F$ ;  $Min := \langle O, O' \rangle$ ;  $Max := \langle F', F \rangle$ ;

**while** ( $Min \neq Max$ ) **do**

    Выбираем случайный элемент  $r \in R$ ;

**if** ( $r \in O$ ) **then**

$Min := CbODown(Min, r)$ ;  $Max := CbODown(Max, r)$ ;

**else**

$Min := CbOUp(Min, r)$ ;  $Max := CbOUp(Max, r)$ ;

**end**

**return**  $Min$ ;

### Алгоритм 2: Спаривающая цепь Маркова

Используя теорему о невозвратных состояниях дискретной цепи Маркова (и лемму Бореля-Кантелли), легко доказать (см., например, [Виноградов, 2018a]), что алгоритм 2 останавливается с вероятностью единица.

### 1.3. Сравнение монотонной и спаривающей цепей Маркова

Сравнение алгоритмов 1 и 2 показывает, что результат спаривающей цепи Маркова можно рассматривать как результат МЦМ для подходящего числа шагов  $T$ .

В [Виноградов, 2018a] была получена точная оценка скорости перемешивания МЦМ порядка  $O(\frac{1}{2} \cdot n \cdot \ln(n))$ .

Метрика **тотальной вариации** между распределениями на конечном множестве (кандидатов)  $\mu, \pi: \{\langle A_j, B_j \rangle: j \in J\} \rightarrow [0,1]$  определяется как половина  $L_1$ -метрики, т. е.  $\|\mu - \pi\|_{TV} = \frac{1}{2} \cdot \sum |\mu_j - \pi_j|$ .

**Теорема 1.** Пусть  $\mu^t$  – распределение перехода за  $t$  шагов для ленивого случайного блуждания по  $n$ -мерной булевой алгебре, а  $\pi$  – равномерное распределение. Тогда  $(\|\mu^t - \pi\|_{TV})^2 \leq \frac{1}{4} \cdot (e^{e^{-c}} - 1)$  при  $t \geq \frac{1}{2} \cdot n \cdot (\ln n + c)$ .

Ранее автор (см., например, [Vinogradov, 2018]) получил точную оценку средней длины траектории спаривающей цепи Маркова.

**Теорема 2.** Для случая  $n$ -мерной булевой алгебры средняя длина траектории спаривающей цепи Маркова равна  $ET = n \cdot \sum_j \frac{1}{j} \approx n \cdot \ln(n) + n \cdot \gamma + \frac{1}{2}$ .

Сравнивая утверждения предыдущих теорем и учитывая наблюдение из начала этого пункта, получаем, что для случая булевой алгебры кандидаты, вычисляемые с помощью СЦМ, имеют распределение, близкое к равномерному. В дальнейшем, мы будем предполагать, что кандидаты имеют равномерное распределение.

## 2. Минимизация эмпирического риска

Алгебраическое машинное обучение по ВКФ-методу соединяет процедуры индуктивного обобщения и предсказания по аналогии.

**Data:** множество обучающих примеров; число  $N$  порождаемых гипотез.

**Result:** подмножество  $S$  гипотез.

**while** ( $i < N$ ) **do**

Породить кандидата  $\langle A, B \rangle$  с помощью алгоритма 2;

$S := S \cup \{\langle A, B \rangle\}$ ;

$i := i + 1$ ;

**end**

### Алгоритм 3: Индуктивное обобщение

Мы устранили проверку «запрета контр-примеров», после которой кандидат, не имеющий контр-примеров, становится ВКФ-гипотезой.

Главным вопросом относительно алгоритма 3 является выбор числа  $N$  порождаемых гипотез.

Используя парадигму, дуальную методу вероятно приближенно-корректного обучения В.Н. Вапника и А.Я. Червоненкиса (см., например, [Вапник и др., 1974]), автор [Виноградов, 2022] доказал следующую теорему.

Теорема 3. Для  $n$  признаков и любых  $\varepsilon > 0$  и  $1 > \delta > 0$  достаточно породить  $N \geq \frac{n \cdot \ln 2 - \ln \delta}{\varepsilon}$  гипотез, чтобы с вероятностью  $1 - \delta$  были правильно предсказаны все  $\varepsilon$ -важные тестовые примеры.

В.Н. Вапник и А.Я. Червоненкис предложили также прямой подход к выбору лучшего классификатора – минимизация эмпирического риска, т.е. числа ошибок, совершаемого классификатором на обучающей выборке.

**Data:** подмножество  $S$  гипотез.

**Result:** значение  $k$  эмпирического риска.

$O$  := обучающие примеры;  $d$  :=  $|O|$ ;  $k$  := 1;

```

for ( $o \in O$ ) do
  for ( $\langle A, B \rangle \in S$ ) do
    if ( $B \subseteq \{o\}$ ) then
       $k := k - 1/d$ ;
    break;
  end
end

```

#### Алгоритм 4: Вычисление эмпирического риска

Для простоты можно попеременно запускать Алгоритмы 3 и 4 до тех пор пока эмпирический риск не перестанет убывать. Но с вычислительной точки зрения выгоднее отмечать родителей для существующих гипотез (абдуктивное объяснение), тогда доля оставшихся неотмеченных обучающих примеров совпадает с эмпирическим риском.

Пусть  $N$  гипотез порождены с помощью спаривающей цепи Маркова, где  $N$  выбрано достаточно большим, чтобы эмпирический риск стал равен  $k = 0$  (это всегда так в случае булевой алгебры без запрета контр-примеров). Так как распределение экспоненциально близко к равномерному, то мы можем порождать гипотезы  $h_j = \langle A_j, B_j \rangle$  (где  $1 \leq j \leq N$ ) из независимых серий испытаний Бернулли ( $\sigma_{j,1}, \dots, \sigma_{j,n}$ ) с вероятностью успеха  $1/2$ :  $A_j = \{o_i : \sigma_{j,i} = 0\}$ ,  $B_j = \{f_i : \sigma_{j,i} = 1\}$ .

**Data:** подмножество  $S$  гипотез.

**Result:** Число  $d$  ошибок предсказаний.

$X$  := тестовые примеры;  $d$  :=  $|X|$ ;

```

for ( $x \in X$ ) do
  for ( $\langle A, B \rangle \in S$ ) do
    if ( $B \subseteq \{x\}$ ) then
       $d := d - 1$ ;
    break;
  end
end

```

### Алгоритм 5: Подсчет ошибок предсказания

Алгоритм 5 соответствует численному варианту рассуждения по аналогии (см. например, [Vinogradov, 2018]), так как предсказываемый положительно тестовый пример содержит общий фрагмент для всех родителей соответствующей гипотезы (и этим фрагментом похож на эти обучающие примеры).

### 3. Асимптотическая оценка переобучения

Как обычно бывает в машинном обучении с учителем, доступные классифицированные примеры делятся на 2 группы: обучающие примеры и тестовые примеры, нужные для проверки качества обученной модели. Тестовые примеры  $X$  образуют обычно лишь малую часть общих данных, но мы допустим, что они составляют половину.

Так как минимальная обучающая выборка для булевой алгебры содержит только коатомы, то возьмем их в качестве тестовых примеров. Точнее, возьмем множество бинарных атрибутов  $F = \{f_1, \dots, f_n\}$  и положим  $x_j = \{f_i \in F: i \neq j\}$  (где  $1 \leq j \leq n$ ).

Для нулевого эмпирического риска  $\eta = 0$ , что всегда можно достичь с вероятностью 1, нам требуется

Лемма 2. Для произвольной функции  $f(n)$  с  $f(n) \rightarrow \infty$  при  $n \rightarrow \infty$  выполняется  $\lim_{n \rightarrow \infty} \mathbf{P}[\eta = 0] = 1$  при  $N \geq \log_2 n + \log_2 f(n)$ .

Док-во.

$$1 \geq \lim_{n \rightarrow \infty} (1 - 2^{-N})^n = \lim_{n \rightarrow \infty} [(1 - 2^{-N})^{2^N}]^{n \cdot 2^{-N}} = \lim_{n \rightarrow \infty} [e^{-1}]^{n \cdot 2^{-N}} \geq \lim_{n \rightarrow \infty} e^{-1/f(n)} = 1.$$

Пусть  $N$  гипотез порождены из независимых серий испытаний Бернулли  $(\sigma_{j,1}, \dots, \sigma_{j,n})$  с вероятностью успеха  $1/2$ :  $h_j = \langle A_j, B_j \rangle$ ,  $A_j = \{o_i: \sigma_{j,i} = 0\}$ ,  $B_j = \{f_i: \sigma_{j,i} = 1\}$  (где  $1 \leq j \leq N$ ).

Обозначим множество гипотез через  $H = \{h_1, h_2, \dots, h_N\}$  и составим соответствующую таблицу

Таблица 1

$XH$	$h_1$	$h_2$	...	$h_N$
$x_1$	0	0	...	0
$\vdots$	$\vdots$	$\vdots$	0	$\vdots$
$x_{\delta_n}$	0	0	...	0
$x_{\delta_{n+1}}$	1	0	...	1
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
$x_n$	0	1	...	1

Здесь единица соответствует тому, что гипотеза вкладывается в пример, т.е. предсказывает его правильно (положительно).



Для доли ошибочных предсказаний  $\theta = \delta$  нужно выбрать  $\delta \cdot n$  строк (это можно сделать  $\binom{n}{\delta \cdot n} = n! / (\delta \cdot n)! ((1 - \delta) \cdot n)!$  способами), в которых должны стоять одни нули, а в остальных хотя бы в одной позиции должны встречаться единицы. Пример из Табл. 1 соответствует такой ситуации с выбором строк  $x_1, \dots, x_{\delta \cdot n}$ .

Задача – оценить  $\mathbf{P}[\theta = \delta] = \binom{n}{\delta \cdot n} \cdot (1 - 2^{-N})^{(n - \delta \cdot n)} \cdot (2^{-N})^{\delta \cdot n}$  при  $n \rightarrow \infty$ .

Для доказательства основного результата нам понадобятся оценка Стирлинга (см., например, [Феллер, 1984]) и верхняя граница энтропии (см., например, [Леонтьев и др., 2019]).

Лемма 3.  $n! = \sqrt{2\pi n} \frac{n^n}{e^n}$  при  $n \rightarrow \infty$ .

Лемма 4.  $-\delta \cdot \ln \delta - (1 - \delta) \cdot \ln(1 - \delta) \leq \ln 2$ .

Теорема 4. Для произвольной функции  $f(n)$  с  $f(n) \rightarrow \infty$  при  $n \rightarrow \infty$  и при  $N \geq \log_2 n + \log_2 f(n)$  выполняется

$$\lim_{n \rightarrow \infty} \mathbf{P}[\theta = \delta] \leq \frac{1}{\sqrt{2\pi\delta(1-\delta)}} \cdot \exp\left\{-\delta \cdot n \ln n + n \cdot \ln 2 - \frac{\ln n}{2}\right\}.$$

Док-во. Очевидно, что  $\mathbf{P}[\theta = \delta] = \binom{n}{\delta \cdot n} \cdot (1 - 2^{-N})^{(n - \delta \cdot n)} \cdot (2^{-N})^{\delta \cdot n}$ . Второй сомножитель не превосходит единицы (и асимптотически равен единице по лемме 2). Третий сомножитель:  $2^{-\delta \cdot n \cdot N} \leq n^{-\delta \cdot n} \cdot f(n)^{-\delta \cdot n} \leq \exp\{-\delta \cdot n \cdot \ln n\}$ . Применим формулу Стирлинга к первому сомножителю. Она после сокращения с помощью верхней границы энтропии дает требуемую асимптотическую оценку.

## Заключение

Основной технический результат работы – верхняя оценка доли ошибочных предсказаний для ВКФ-метода в простом случае, когда обучающие и тестовые примеры представляют собой коатомы в булевой алгебре, а контр-примеры не учитываются.

Аргумент за выбор коатомов в качестве тестовых примеров – порождение гипотез из минимальной обучающей выборки для булевой алгебры, где ее элементы тоже коатомы. Если выбирать произвольные элементы булевой алгебры в качестве тестовых примеров, а гипотезы порождать из коатомов, то теорема 4 перестает быть верной. Но такой подход противоречит идеи тестирования на части классифицированных примерах, когда эти примеры случайным образом разделяются на обучающую и тестовую выборку, т.е. представляют собой однородные данные.

Однако главное значение работы – установление того факта, что многие идеи (абдукция и аналогия) ДСМ-метода автоматического порождения гипотез [Финн и др., 2009] имеют численные варианты для ВКФ-метода алгебраического машинного обучения, которые совпадают с мето-

дом минимизации эмпирического риска и расчета числа ошибок при предсказании тестовых примеров, ранее предложенного В.Н. Вапником и А.Я. Червоненкисом.

**Благодарности.** Автор благодарит своих коллег по ВЦ им. А.А. Дородницына ФИЦ ИУ РАН за поддержку и полезные дискуссии. Особая признательность выражается Л.А. Якимовой за совместную работу, обсуждения и поддержку.

### Список литературы

- [Vinogradov, 2018] Vinogradov D.V. Machine Learning Based on Similarity Operation // Communication in Computer and Information Science. – 2018. – Vol. 934. – P. 46-59.
- [Виноградов, 2018a] Виноградов Д.В. Вероятностно-комбинаторный формальный метод обучения, основанный на теории решеток: дис. ... д-ра физ.-мат. наук 05.13.17. – М.: ФИЦ ИУ РАН, 2018.
- [Виноградов, 2018b] Виноградов Д.В. Скорость сходимости к пределу вероятности порождения случайного сходства при наличии контр-примеров // Научная и техническая информация. Сер. 2. – 2018. – № 2. – С. 21-24.
- [Виноградов, 2022] Виноградов Д.В. Алгебраическое машинное обучение: упор на эффективность // Автоматика и телемеханика. – 2022. – № 6. – С. 5-23.
- [Вапник и др., 1974] Вапник В.Н., Червоненкис А.Я. Теория распознавания образов (статистические проблемы обучения). – М.: Наука, 1974.
- [Финн и др., 2009] Финн В.К., Аншаков О.М. (ред.) ДСМ-метод автоматического порождения гипотез: логические и эпистемологические основания. – М.: URSS, 2009.
- [Кузнецов, 1993] Кузнецов С.О. Быстрый алгоритм построения всех пересечений объектов из нижней полурешетки // Научная и техническая информация. Сер. 2. – 1993. – № 1. – С. 17-20.
- [Феллер, 1984] Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х т. Т. 1: пер. с англ. – М.: Мир, 1984.
- [Леонтьев и др., 2019] Леонтьев В.К., Гордеев Э.Н. Комбинаторные аспекты теории информации. – М.: МФТИ, 2019.

## БЫСТРЫЙ ПОИСК ОПТИМАЛЬНЫХ КОРОТКИХ КЛАССИФИКАТОРОВ

Е.О. Дудырев (*eo.dudyrev@hse.ru*)

С.О. Кузнецов (*skuznetsov@hse.ru*)

Национальный исследовательский университет  
«Высшая школа экономики», Москва

Исследования в области объяснимого искусственного интеллекта (Explainable AI, XAI) показывают, что модель должна быть маленькой, чтобы быть понятной человеку. Ограничение на размер модели резко сокращает пространство возможных решений. При этом многие модели обучения правилам по-прежнему полагаются на жадные алгоритмы для создания ансамблей деревьев решений. В этой статье обсуждаются математические и инженерные методы, основанные на идеях Анализа Формальных Понятий, для эффективного поиска оптимальных коротких бинарных классификаторов, состоящих не более чем из трех бинарных атрибутов и являющимися оптимальными относительно F1 меры (среднего гармонического между мерами точности и полноты).

**Ключевые слова:** бинарная классификация, объяснимое машинное обучение, машинное обучение с учителем, Анализ Формальных Понятий.

### Введение

Исследования в области объяснимого искусственного интеллекта (Explainable AI, XAI) показывают, что человеку понятны лишь маленькие модели машинного обучения. Но строгие критерии малого размера модели до сих пор не разработаны. В том числе, так как считается, что малые модели не могут дать хорошего качества предсказаний. В данной работе мы предлагаем алгоритм, который позволяет находить малые модели с качеством предсказаний, сравнимым с передовыми решениями в данной области.

Современные модели обучения, такие как градиентный бустинг на деревьях решений (Friedman, 2001) или случайный лес (Breiman, 2001), повсеместно признаны основными современными решениями для задач бинарной классификации. Однако, даже с использованием техник объясни-

мого машинного обучения (ХАИ) (Molnar, 2020) (Barredo Arrieta, et al., 2020), такие модели слишком сложны для анализа и использования в сценариях, требующих доверия пользователя. Это является причиной растущей тенденции отказа от сложных моделей черного ящика в пользу небольших объяснимых моделей (Rudin, 2019) (Pianykh, et al., 2022).

Ограничение размера модели не только делает такие модели легко объяснимыми, но также резко ограничивает пространство всех возможных моделей. Таким образом, вместо поиска модели, локально оптимальной по качеству, становится возможным поиск глобального оптимума. В данной статье рассматриваются модели, оперирующие только тремя бинарными (двоичными) атрибутами. Чтобы найти оптимальную модель, удовлетворяющую этим условиям, необходимо перебрать все комбинации от одного до трех заданных бинарных атрибутов. Однако с увеличением количества атрибутов такой алгоритм грубой силы страдает от комбинаторного взрыва. В этой статье рассматриваются как математические, так и инженерные методы, основанные на анализе формальных понятий (АФП) (Ganter, et al., 1999), которые делают этот алгоритм грубой силы более эффективным.

Задача построения простых логических моделей решалась во многих областях предыдущих исследований. В начале 1960-х работа над формальной логикой привела к появлению алгоритмов логического программирования и обучающихся правил (Bongard, 1970), (Michalski, 1973). Модели на правилах, включая такие алгоритмы, как Scope-Rules, RuleFit (Friedman, et al., 2008) и другие, часто полагаются на жадные подходы для извлечения коротких правил из более сложных моделей (таких как большие деревья решений). В отличие от этих жадных подходов, приводящих к локально оптимальным моделям, в этой статье рассматривается проблема поиска глобально оптимальной модели заданного размера.

Другая возможная причина для разработки и использования коротких моделей заключается в том, что они являются хорошим подспорьем для оценки качества больших моделей черного ящика. В самом деле, если короткая модель выдает такое же качество предсказания, как и большой черный ящик, то использование последнего – бессмысленно.

Обширные исследования показывают, что человек может оперировать посылками (идеями), содержа в голове одновременно не более четырех (плюс минус одну) базовых посылок (Cowan, 2001) (Halford, et al., 2005).

В этой статье из-за ограничений алгоритмической сложности мы сосредоточимся на поиске правил с посылками, имеющими не более трех атрибутов, по двум причинам. Во-первых, так как это наш первый подход к проблеме, мы должны попытаться решить ее самый простой вариант. Во-вторых, выбор числа три обоснован когнитивными исследованиями.

Таким образом, короткие правила, состоящие из более чем трех атрибутов, представляют собой особый класс объяснимых моделей машинного обучения.

Структура статьи следующая. В разделе 1 представлены теоретические основы, использованные на протяжении всего работы. Раздел 2 описывает математические приемы оптимизации алгоритма путем минимизации количества требуемых операций. В дополнение к этому в Разделе 3 обсуждаются инженерные подходы к оптимизации алгоритма за счет максимизации скорости вычислений. Раздел 4 объединяет все обсуждаемые методы в один алгоритм. В разделе 5 представлены экспериментальные результаты этого алгоритма.

## 1. Теоретические основания

В этом подразделе вводятся определения, которые мы используем на протяжении всей статьи. Во-первых, мы представляем основные термины Анализа Формальных Понятий для описания моделей правил. Во-вторых, мы описываем пространство посылок, в котором находятся рассматриваемые в статье модели машинного обучения. В-третьих, мы опишем основные темы бинарной классификации на языке нотации АФП.

Анализ Формальных Понятий. Формальный контекст  $K$  описывает набор данных, на котором строится модель. Он представлен в виде тройки  $K = (G, M, I)$ , где  $G$  – набор объектов (строк в наборе данных),  $M$  – набор атрибутов (столбцов в наборе данных),  $I \subseteq G \times M$  представляет отношения между объектами и атрибутами.

Оператор замыкания ( $'$ ) сопоставляет подмножество объектов  $A \subseteq G$  подмножеству атрибутов  $B$  таким образом, что объекты из  $A$  «описываются» атрибутами из  $B$  и наоборот:

$$A' = \{m \in M \mid \forall g \in A: gIm\} \quad B' = \{g \in G \mid \forall m \in B: gIm\}. \quad (1)$$

Для заданного подмножества атрибутов  $B \subseteq M$  подмножество объектов  $A = B'$  называется объёмом  $B$ . Аналогично, подмножество атрибутов  $B = A'$  называется содержанием  $A$ .

Посылки. В этом подразделе мы определяем посылку как комбинацию атрибутов формального контекста, соединённых операциями конъюнкции, дизъюнкции и отрицания. Понятие посылки необходимо для следующих подразделов.

Пространство посылок  $\mathbb{P}$  – это множество всех комбинаций атрибутов  $M$ , построенных с помощью операций конъюнкции  $\wedge$ , дизъюнкции  $\vee$  и отрицания  $\neg$ :

$P$  – это множество, такое что:

$$1) M \subseteq P$$

$$2) \forall p, q \in P: p \wedge q, p \vee q, \neg p \in P. \quad (2)$$

Каждой посылке  $p \in \mathbb{P}$  соответствует подмножество объектов  $p' \subseteq G$ , называемое объёмом посылки. Объёмы операций конъюнкции, дизъюнкции и отрицания определяются следующим образом:

$$(p \wedge q)' = p' \cap q', (p \vee q)' = p' \cup q', (\neg p)' = G \setminus p'. \quad (3)$$

В настоящей работе особенно интересны посылки, состоящие не более чем из трех атрибутов.

В общем случае набор посылок  $P_i$ , построенный из  $i$  атрибутов, можно описать следующим образом:

$$P_1 = M \cup \{\neg m \mid m \in M\}$$

$$P_{i,i>1} = \bigcup_{j=1}^{\lfloor i/2 \rfloor} \{p \wedge q, p \vee q, \neg(p \wedge q), \neg(p \vee q) \mid p \in P_j, q \in P_{i-j}\}. \quad (4)$$

Объединяя множества посылок  $P_i$  для каждого натурального числа  $i$ , мы получаем исходное пространство посылок  $\mathbb{P} = \bigcup_{i=1}^{\infty} P_i$ . В дальнейшем мы будем говорить, что посылка  $p \in \mathbb{P}$  имеет размер  $i$ , если она принадлежит  $P_i$ .

**Бинарная классификация.** Бинарная классификация – это задача машинного обучения, когда модель просят предсказать, принадлежит ли объект  $g$  к «положительному» или «отрицательному» классу, учитывая описание объекта (заданное подмножеством атрибутов). Модель получена на основе предоставленного обучающего контекста (набора данных)  $K = (G, M, I)$  с предопределенными положительными  $G_+ \in G$  и отрицательными  $G_- = G \setminus G_+$  объектами. Первый шаг к построению хорошего бинарного классификатора — найти модель, работающую с набором атрибутов  $M$ , которая эффективно отделяет положительные объекты  $G_+$  и отрицательные объекты  $G_-$ . После этого модель можно применить к тестовому контексту  $K_{\text{test}} = (G_{\text{test}}, M, I_{\text{test}})$  для прогнозирования его доселе неизвестных положительных и отрицательных объектов.

В данной работе исследуются бинарные классификаторы вида «если посылка  $p \in \mathbb{P}$  верна, то объект  $g$  прогнозируется положительным, иначе объект  $g$  прогнозируется отрицательным».

Качество предсказания посылки  $p \in \mathbb{P}$  на обучающем наборе данных измеряется путем сравнения заданных наборов положительных  $G_+$  и отрицательных  $G_-$  объектов с наборами положительных  $G_{p+}$  и отрицательных объектов  $G_{p-}$  прогнозируемых посылкой  $p$ . Отметим, что множество  $G_{p+}$  в точности соответствует посылке  $p$ :  $G_{p+} = p'$ .

$$\begin{aligned}
TP_p &= G_+ \cap G_{p_+} = G_+ \cap p' & FP_p &= G_- \cap G_{p_+} = G_- \cap p' \\
FN_p &= G_+ \cap G_{p_-} = G_+ \setminus p' & TN_p &= G_- \cap G_{p_-} = G_- \setminus p'.
\end{aligned} \tag{5}$$

Для краткости будем использовать строчные буквы для обозначения мощностей так называемых истинно положительных предсказаний  $TP_p$ , ложно положительных предсказаний  $FP_p$ , ложно отрицательных предсказаний  $FN_p$  и истинно отрицательных предсказаний  $TN_p$ . Точно так же мы используем  $y_+$ ,  $y_-$  для обозначения мощностей множества положительных объектов  $G_+$  и множества отрицательных объектов  $G_-$  соответственно:

$$\begin{aligned}
tp_p &= |TP_p| & fp_p &= |FP_p| & fn_p &= |FN_p| & tn_p &= |TN_p| \\
y_+ &= |G_+| & y_- &= |G_-|.
\end{aligned} \tag{6}$$

Одними из наиболее широко используемых показателей качества для бинарных классификаций являются точность (*prec*), полнота (*rec*) и их среднее гармоническое, называемое F1 мерой (*F1*). Их определения следующие:

$$prec(p) = \frac{tp_p}{|p'|}, \quad rec(p) = \frac{tp_p}{y_+}, \quad F1(p) = 2 \frac{prec(p) * rec(p)}{prec(p) + rec(p)}. \tag{7}$$

В этой статье основное внимание уделяется нахождению посылки  $p^*$  размера не больше чем 3, имеющей максимальную значение *F1* меры:

$$p^* = arg \max_{p \in \cup_{i=1}^3 P_i} F1(p). \tag{8}$$

Представленные техники могут быть легко адаптированы для других показателей качества.

## 2. Минимизация количества сравнений

**Оптимизация F1 меры.** Определение показателя F1 как гармонического среднего значения точности и полноты делает неясными возможные стратегии оптимизации. Этот подраздел упрощает задачу максимизации балла F1 за счет максимизации количества истинно положительных и истинно отрицательных результатов.

F1 мера  $F1(p)$  комонотонна оценке Жаккара  $J(p)$  (обозначается  $\sim$ ), где последняя представляет собой коэффициент подобия Жаккара между множеством положительных объектов  $G_+$  и множеством объектов предсказанных положительными  $G_{p_+} = p'$ :

$$F1(p) \sim J(p) = \frac{|G_+ \cap p'|}{|G_+ \cup p'|}. \tag{9}$$

Поскольку F1 мера  $F1(p)$  комонотонна по отношению к показателю Жаккара  $J(p)$ , то задачу оптимизации F1 меры можно рассматривать как задачу оптимизации дроби  $tp_p/|G|-tn_p$ , то есть задачу максимизации истинно положительных и истинно отрицательных предсказаний:

$$\arg \max_{p \in \mathbb{P}} F1(p) = \arg \max_{p \in \mathbb{P}} J(p) = \arg \max_{p \in \mathbb{P}} \frac{tp_p}{|G| - tn_p}. \quad (10)$$

**Влияние логических операций на оценку Жаккара.** В этом подразделе обсуждается, как операции конъюнкции и дизъюнкции влияют на оценку Жаккара. То есть, учитывая оценку Жаккара двух посылок  $p, q \in \mathbb{P}$ , можем ли мы ожидать, что у посылок  $p \wedge q, p \vee q$  будут более высокие или более низкие значения оценки Жаккара?

Во-первых, давайте выразим истинные положительные и истинные отрицательные стороны посылок, построенных конъюнкцией и дизъюнкцией, с истинными положительными и истинными отрицательными сторонами исходных посылок:

$$\begin{aligned} TP_{p \wedge q} &= G_+ \cap (p' \cap q') = TP_p \cap TP_q & (11) \\ TP_{p \vee q} &= G_+ \cap (p' \cup q') = TP_p \cup TP_q \\ TN_{p \wedge q} &= G_- \setminus (p' \cap q') = TN_p \cup TN_q \\ TN_{p \vee q} &= G_- \setminus (p' \cup q') = TN_p \cap TN_q \end{aligned}$$

Следовательно, операция конъюнкции сужает множество истинных положительных предсказаний при расширении набора истинно отрицательных. Напротив, операция дизъюнкции расширяет множество истинно положительных предсказаний и сужает множество истинно отрицательных.

Теперь мы выводим уравнения для инфимума и супремума для мощностей истинно положительных и истинно отрицательных предсказаний. Для этого мы используем две идеи из уравнений (15). Во-первых, мы используем установленные ограничения мощности для операций пересечения и объединения. Во-вторых, поскольку число истинно положительных предсказаний  $tp$  ограничено числом положительных объектов  $y_+$ , то для двух посылок  $p, q \in \mathbb{P}$ , если сумма  $tp_p, tp_q$  превышает  $y_+$ , то соответствующие истинно положительные предсказания должны иметь не менее  $tp_p + tp_q - y_+$  общих объектов (аналогичные выводы можно сделать для числа истинно отрицательных предсказаний  $tn$  и числа отрицательных объектов  $y_-$ ).

$$\begin{aligned} \max(tp_p + tp_q - y_+, 0) &\leq \mathbf{tp}_{p \wedge q} \leq \min(tp_p, tp_q) & (12) \\ \max(tn_p, tn_q) &\leq \mathbf{tn}_{p \wedge q} \leq \min(tn_p + tn_q, y_-) \\ \max(tp_p, tp_q) &\leq \mathbf{tp}_{p \vee q} \leq \min(tp_p + tp_q, y_+) \\ \max(tn_p + tn_q - y_-, 0) &\leq \mathbf{tn}_{p \vee q} \leq \min(tn_p, tn_q) \end{aligned}$$



Таким образом, границы оценки Жаккара следующие:

$$\frac{\max(tp_p + tp_q - y_+, 0)}{|G| - \max(tn_p, tn_q)} \leq J(p \wedge q) \leq \frac{\min(tp_p, tp_q)}{|G| - \min(tn_p + tn_q, y_-)} \quad (13)$$

$$\frac{\max(tp_p, tp_q)}{|G| - \max(tn_p + tp_q - y_-, 0)} \leq J(p \vee q) \leq \frac{\min(tp_p + tp_q, y_+)}{|G| - \min(tn_p, tn_q)}$$

Рис. 1 визуализирует границы на плоскости TruePositive-TrueNegative (истинно положительных и истинно отрицательных предсказаний) для абстрактного формального контекста, где 40% объектов являются положительными. Для максимизации показателя Жаккара, показанного увеличением градиента серого цвета, нам необходимо достичь правого верхнего угла плоскости. Однако операции конъюнкции и дизъюнкции над посылками  $p, q \in \mathbb{P}$  могут лишь «переместить» посылки в левый верхний ( $p \vee q$ ) и правый нижний ( $p \wedge q$ ) углы.

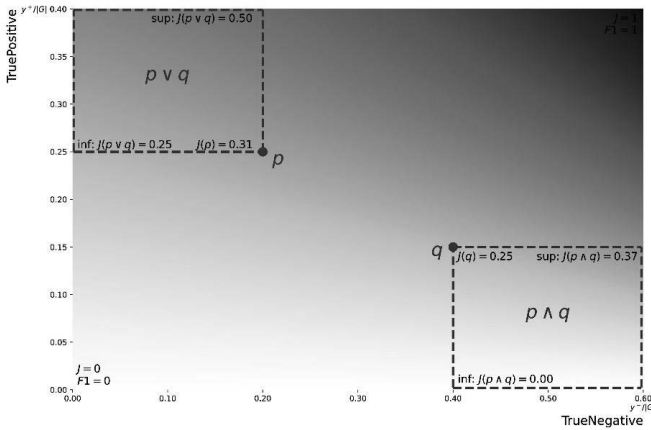


Рис. 1. Оценка Жаккара операции конъюнкции и дизъюнкции над абстрактными предпосылками  $p, q$  абстрактного формального контекста. Интенсивность серого цвета показывает значение показателя Жаккара для конкретной точки на плоскости

Границы, полученные для операций конъюнкции и дизъюнкции, расплывчатые: они не говорят, будет ли вновь сформированная посылка иметь более высокий или более низкий показатель Жаккара. Более точная оценка возможна только при пересечении объемов посылок  $p, q$ . Однако в этом случае можно сразу же вычислить результирующую оценку Жаккара, а не оперировать оценками.

Учитывая пороговое значение  $\theta \in \mathbb{R}$  и посылку  $p \in \mathbb{P}$ , можно определить, соответствует ли оценка Жаккара любой посылке  $p \wedge q$ ,  $p \vee q \in \mathbb{P}$  не превысит порог  $\theta$  с помощью следующих неравенств:

$$tp_p \leq y_+ \theta \rightarrow J(p \wedge q) \leq \theta, \quad \forall q \in \mathbb{P} \quad (14)$$

$$tn_p \leq |G| - \frac{y_+}{\theta} \rightarrow J(p \vee q) \leq \theta, \quad \forall p \in \mathbb{P}.$$

### 3. Максимизация скорости сравнений

Математические приемы помогают минимизировать количество правил сравнения. Однако количество таких правил по-прежнему велико. В этом разделе основное внимание уделяется инженерным приемам, позволяющим ускорить каждое сравнение.

**Объемы и битовые списки.** В этом подразделе мы используем два важных факта об объемах посылок: (i) разным посылкам может соответствовать один и тот же объем  $\exists p, q \in \mathbb{P}: p \neq q, p' = q'$ , (ii) любая мера качества предсказания посылки  $p$  рассчитывается на основе объема посылки  $p'$  (см. уравнения (5)).

Разные посылки  $p, q \in \mathbb{P}$ ,  $p \neq q$  могут соответствовать одному и тому же объёму  $p' = q' \subseteq G$  по многим причинам. Во-первых, посылки могут быть логически эквивалентны: т.е.  $\neg(p \wedge q)' = (\neg p \vee \neg q)'$  по законам Де Моргана. Во-вторых, если одна посылка менее общая, чем другая  $p' \subset q'$ , то их конъюнкция будет соответствовать объёму  $p'$ , а их дизъюнкция будет соответствовать объёму  $q'$ . Наконец, разные посылки могут соответствовать одним и тем же объёмам из-за специфических характеристик формального контекста  $K$ : например может случиться так, что объём конъюнкции двух посылок  $p, q \in \mathbb{P}$  будет равен объёму третьей посылки  $r \in \mathbb{P} \setminus \{p, q\}$ :  $(p \wedge q)' = r'$ .

Поскольку вычисление мер качества предсказания зависит от объёмов предпосылок, то многие различные предпосылки, соответствующие одному и тому же объёму, будут иметь одинаковое качество предсказания (применительно к контексту  $K$ ).

Таким образом, мы предлагаем искать оптимальный объём, а не оптимальную посылку. Чтобы формализовать эту идею, определим объём множеств  $E_i$ :

$$E_1 = \{p' \mid \forall p \in P_1\} = \bigcup_{\{m \in M\}} \{m', (\neg m)'\} \quad (15)$$

$$E_{i, i > 1} = \bigcup_{j=1}^{i//2} \{a \wedge b, a \vee b \mid a \in E_j, b \in E_{i-j}\} \setminus \bigcup_{j=1}^{i-1} E_j$$

Предлагаемое определение множеств объёмов гарантирует, что любой объём  $e \in E_i$  порождается посылкой размера не менее  $i$ :

$$\forall i, j \in N, j < i, e \in E_i: \exists p \in P_i: p' = e, \nexists q \in P_j: q' = e. \quad (16)$$

Таким образом, задача оптимизации становится следующей:

$$e^* = \arg \max_{e \in \cup_{i=1}^3 E_i} F1(e), \quad (17)$$

где функция оценки F1 меры немного изменена, чтобы использовать в качестве параметра объём, а не посылку.

И последнее, но не менее важное: объём, являющийся подмножеством объектов  $G$ , может быть представлен и сохранен в компьютере как битовая маска (набор битов) длины  $|G|$ , где каждый бит представляет, находится ли соответствующий объект в объёме или нет. Операции конъюнкции и дизъюнкции становятся операциями над битовыми масками, которые являются наиболее эффективными операциями, выполняемыми современными компьютерами с двоичным кодом. Таким образом, использование объёмов вместо посылок не только сокращает количество сравнений, но и значительно ускоряет каждое из сравнений.

**Операции над списками.** В этом подразделе описывается прием, который хорошо известен среди специалистов по науке о данных, однако мы должны описать его для полного раскрытия темы.

Неравенства, представленные в формулах 14, позволяют обойтись без обработки многих конъюнкций  $p \wedge q$  и дизъюнкций  $p \vee q$  на основе характеристик исходных посылок  $p, q \in \mathbb{P}$  и порога качества предсказания  $\theta$ . Однако эти характеристики еще необходимо вычислить. Для вычисления этих числовых характеристик наиболее эффективным образом мы используем пакет Numpy (Harris, et al., 2020) для Python. Пакет специально разработан для работы с большими объемами числовых данных за счет использования кода C++.

#### 4. Предлагаемый алгоритм

Вот псевдокод алгоритма поиска наилучших  $k$  посылок размера не больше чем 3:

- Шаг 1. Найдём все объёмы размера 1, которые не потеряют в качестве после конъюнкции, дизъюнкции:
  - a. Посчитаем все объёмы  $E_j$ ;
  - b. Найдём порог качества  $\theta$  как минимальное качество  $k$  лучших объёмов из  $E_j$ ;
  - c. Оставим из всех объёмов  $E_j$  только те, которые удовлетворяют неравенствам (14).

- Шаг 2. Найдём все объёмы размера 2, которые не потеряют в качестве после конъюнкции и дизъюнкции:
  - a. Рассчитаем все объёмы  $E_2$ , основанные на отфильтрованном множестве  $E_1$ . В процессе сохраняя информацию о том, какая пара объёмов и операция породили каждый из новых объёмов;
  - b. Найдём порог качества  $\theta$  как минимальное качество  $k$  лучших объёмов из  $E_1$  и  $E_2$ ;
  - c. Оставим из всех объёмов  $E_1$  и  $E_2$  только те, которые удовлетворяют неравенствам (14).
- Шаг 3. Найдём только  $k$  лучших объёмов размера 3: Для каждой пары объёмов  $e_1, e_2$  из отфильтрованных  $E_1, E_2$ :
  - a. Если любой из экстенгов  $e_1, e_2$  удовлетворяет обоим неравенствам (14), то переходим к следующей паре; в противном случае:
  - b. Вычислим и измерим качество предсказания конъюнкции  $e_1 \cap e_2$ ;
  - c. Вычислим и измерим качество предсказания дизъюнкции  $e_1 \cup e_2$ ;
  - d. При необходимости обновим  $\theta$ .
- Шаг 4. Восстановите посылки, соответствующие лучшим  $k$  объёмам используя сохранённую информацию об объёмах и операциях.

Временная сложность этого алгоритма равна  $O(|M|^3)$ , где  $M$  – набор атрибутов в формальном контексте  $K$ . С асимптотической точки зрения это та же временная сложность, что и у алгоритма грубой силы для проверки всех посылок размером не более чем три. Однако использование объёмов, а также уменьшение количества комбинаций позволяет минимизировать практическое время обработки алгоритма.

## 5. Эксперименты

В этом разделе предложенный алгоритм применяется на практике. Во-первых, мы изучаем статистику количества сравнений, которые должен сделать алгоритм. Во-вторых, мы сравниваем качество предсказания коротких моделей с качеством моделей черного ящика, чтобы показать, что есть случаи, когда первые работают так же эффективно, как вторые.

Алгоритм запускается на реальном наборе данных Myocard (Golovenkin, et al., 2020) из репозитория UCI. Набор данных содержит 1700 объектов и 124 атрибута. Задача набора данных Myocard состоит в том, чтобы предсказать, будет ли у пациента в больнице хроническая сердечная недостаточность, на основе его данных. Нам не удалось запустить алгоритм за короткое время (т.е. часы и дни) на больших наборах данных из-за комбинаторного взрыва. Однако мы считаем, что набор данных Myocard достаточно велик для тестирования алгоритма.

**Количество сравнений.** Табл. 1 показывает, что использование границ из неравенств (14) не отфильтровывает многие комбинации объёмов. Например, для посылки размером 3 границы фильтруют только 21,5% комбинаций объёмов (от  $1,49e+09$  до  $1,17e+09$ ). Хотя такой процент лучше, чем ничего, он все же требует тестирования 1,17 миллиарда комбинаций объёмов. Тем не менее, использование битовых массивов позволяет протестировать 1,17 миллиарда комбинаций экстенгов всего за 1 час на ноутбуке с 8 ГБ ОЗУ.

Таблица 1

**Статистика по количеству итераций алгоритма**

Размер посылки	1	2	3
# посылок	1752	$3.07e+06$	$1.08e+10$
# объёмов	1644	$2.53e+06$	$1.49e+09$
# объёмов для проверки	1644	$2.53e+06$	$1.17e+09$
# новых объёмов	1644	$9.44e+05$	$9.55e+08$
# сохраняемых объёмов	1644	$9.44e+05$	259
время вычисления	7.08 мс	6.55 с	1.06 ч

Описание табл. 1:

- # посылок: количество посылок заданного размера; Это комбинаторно вычисленное максимальное количество итераций алгоритма;
- # объёмов: количество комбинаций объёмов, в результате которых получается посылка заданного размера;
- # объёмов для проверки: количество комбинаций объёмов, которые могут привести к хорошему качеству предсказания (отфильтровано по неравенствам (14));
- # новых объёмов: количество новых объёмов, созданных в результате тестирования комбинаций объёмов;
- # сохраняемых объёмов: количество объёмов, которые необходимо сохранить в памяти. Для размера 1 и 2 мы сохраняем все объёмы, для размера 3 оставляем только объёмы с высоким качеством предсказания;
- время вычисления: время, затраченное на обработку всех комбинаций для данного размера посылки.

**Качество предсказаний коротких правил.** Простота моделей коротких правил позволяет полностью привести некоторые из полученных моделей в статье. В следующем списке представлены лучшие короткие модели (по показателю F1, определенному стандартным способом для метода бинарной классификации), полученные в наборе данных Myocard:

- посылка размера 1: F1 мера = 0,401426;  
посылка: имеются данные об использовании обезболивающих средств в реанимационном отделении на третьи сутки госпитального периода;
- посылка размера 2: F1 мера = 0,448819;  
посылка: (у человека была хроническая сердечная недостаточность) ИЛИ (имеет в анамнезе сахарный диабет);
- посылка размера 3: F1 мера = 0,473786;  
посылка: (есть данные об употреблении опиоидных препаратов в отделении интенсивной терапии на третий день госпитального периода) И (Была хроническая сердечная недостаточность) ИЛИ (Возраст  $\geq 66$ );
- подель XGBoost: F1 мера = 0,464000;  
подель содержит 100 деревьев решений максимальной глубины 6;
- подель CatBoost: F1 мера = 0,434783;  
подель содержит 1000 деревьев решений глубины 6.

Мы также можем показать все полученные короткие модели правил в пространстве TruePositive-TrueNegative.

На рис. 2 показано качество предсказания полученных коротких моделей по шкале TruePositive-TrueNegative. Следует отметить, что точки, соответствующие сложным моделям градиентного бустинга лежат недалеко от точек коротких моделей. Таким образом, использовать сложные черные ящики на наборе данных Myocard нецелесообразно, так как простые короткие модели обеспечивают такое же качество предсказания.

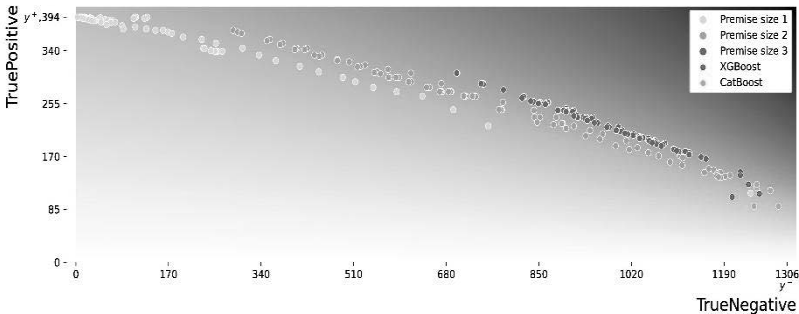


Рис. 2. Качество предсказания коротких моделей в пространстве TruePositive-TrueNegative. Качество прогнозирования моделей XGBoost и CatBoost по умолчанию представлено в качестве эталона

## Заключение

В этой статье мы представили предварительные результаты по поиску оптимального правила с посылкой, состоящей не более чем из трех бинарных признаков. Мы описали задачу оптимизации F1 меры с точки зрения истинно положительных и истинно отрицательных прогнозов. Мы вычислили верхнюю и нижнюю границы коэффициентов Жаккара для посылок, полученных с помощью операций конъюнкции и дизъюнкции. Мы также рассмотрели технику перебора объёмов посылок на основе идей из АФП, чтобы минимизировать время выполнения вычислений.

В следующих исследованиях мы планируем разработать более точные нижние и верхние границы оценки Жаккара для посылок, построенных с помощью операций конъюнкции и дизъюнкции. Мы также планируем обсудить другие логические операции, которые повысят качество прогнозирования правил при неизменном количестве используемых признаков.

**Благодарности.** Статья подготовлена в рамках Программы фундаментальных исследований НИУ ВШЭ, РФ.

## Список литературы

- [**Angelino Elaine и др.**] Learning Certifiably Optimal Rule Lists // Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – New York, NY, USA: Association for Computing Machinery, 2017. – P. 35-44.
- [**Barredo Arrieta Alejandro и др.**] Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI // Information Fusion. – 2020. – P. 82-115.
- [**Bongard Mikhail Moiseevich**] Pattern Recognition. – Rochelle Park, N.J.: Hayden Book Co., Spartan Books, 1970.
- [**Breiman Leo**] Random Forests // Machine Learning. – 2001.
- [**Cowan N.**] The magical number 4 in short-term memory: A reconsideration of mental storage capacity // Behavioral and Brain Sciences. – 2001. – 24 (1). – P. 87-114.
- [**Friedman Jerome**] Greedy Function Approximation: A Gradient Boosting Machine // Annals of Statistics. – 10 2001. – Vol. 29. – P. 1189-1232.
- [**Friedman Jerome H. и Popescu Bogdan E.**] Predictive learning via rule ensembles // The annals of applied statistics. – Institute of Mathematical Statistics, 2008. – 3. – Vol. 2. – P. 916-954.
- [**Ganter B. и Wille R.**] Formal Concept Analysis: Mathematical Foundations. – Springer Berlin Heidelberg, 1999.
- [**Golovenkin Sergey E. и др.**] Trajectories, bifurcations, and pseudo-time in large clinical datasets: applications to myocardial infarction and diabetes data // GigaScience. – November 2020.

- [**Halford G.S и др.**] How Many Variables Can Humans Process? // Psychological Science. – 2005. – 16 (1). – P. 70-76.
- [**Harris Charles R. и др.**] Array programming with {NumPy} // Nature. – Springer Science and Business Media {LLC}. – 9 2020. – 7825. – Vol. 585. – P. 357-362.
- [**Michalski Ryszard S.**] Discovering classification rules using variable-valued logic system VL1 // the Third International Joint Conference on Artificial Intelligence. – Stanford, California. – 1973. – P. 162-172.
- [**Miller George A.**] The magical number seven, plus or minus two: Some limits on our capacity for processing information / Psychological review. – 1956. – 63 (2). – P. 81.
- [**Molnar Christoph**] Interpretable machine learning. – Lulu. com, 2020.
- [**Pianykh Oleg и др.**] Human Knowledge Models: Learning Applied Knowledge from the Data // Unpublished. – 2022.
- [**Rudin Cynthia**] Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead // Nature Machine Intelligence. – 2019. – P. 206-215.



## ТРИ ВОПРОСА (НА ПОНИМАНИЕ), АДРЕСОВАННЫЕ «ТОВАРИЩАМ ПО ПАРТИИ»

М.И. Забежайло (*m.zabzhailo@yandex.ru*)  
ФИЦ «Информатика и управление» РАН, Москва

Обсуждаются представления о «качестве» результатов интеллектуального анализа данных (ИАД). Рассматриваются «качество» обучающей выборки прецедентов, используемого языка представления знаний и собственно «инструментария» ИАД. Сформулирован ряд вопросов, приемлемые ответы на которые позволят повысить результативность и «качество» ИАД.

**Ключевые слова:** искусственный интеллект, интеллектуальный анализ данных, исследования и разработки, оценки качества.

В последние годы путаница в толкованиях термина Искусственный Интеллект (ИИ) стала обыденным явлением. На конференции КИИ-2021 об этом отдельно говорилось в сообщении [Финн и др. 2021]. Нередко приходится сталкиваться со своего рода «бурями в стакане воды», причина которых – те или иные смысловые дефекты определения ряда ИИ-терминов.

Между тем, можно показать, что не всё обстоит абсолютно безупречным (*«не оспариваемым в честном споре»*) также и с некоторыми рабочими понятиями, активно используемыми ИИ-профессионалами. Каждому из нас хорошо знакомы понятия *обучающая выборка, язык представления знаний (ЯПЗ), инструмент* (метод, алгоритмическая конструкция, ...) *интеллектуального анализа данных (ИАД)* и т.п. А можно ли оценить «качество» выборки прецедентов, ЯПЗ а также процедурного инструментария анализа данных и поддержки принятия решений (АД и ППР), выбираемых при решении конкретной проблемы? Например, что такое *«хорошая»* обучающая выборка, и чем она отличается от *«плохой»*? Таким образом, по-видимому, вполне уместен адресованный всему сообществу ИИ-специалистов вопрос: так чем же мы оперируем, вообще говоря, по умолчанию предполагая, что использование названных выше понятий носит вполне корректный характер (в частности, не несет тех или иных смысловых дефектов)?

## 1. Общие представления об ИАД и его «качестве»

Конкретный вариант реализации интеллектуального анализа данных можно охарактеризовать упорядоченной тройкой:

$$\Sigma = \langle DS, ЯПЗ, ИАД-М \rangle,$$

где

**DS** – набор анализируемых данных (обучающая выборка, data set, ...),

**ЯПЗ** – конкретный вариант ЯПЗ об изучаемом эффекте/явлении, используемый для «кодирования» данных из анализируемого набора **DS**,

**ИАД-М** – конкретный вариант процедурных средств, обеспечивающих восстановление скрытых в **DS** эмпирических зависимостей.

В каждом конкретном случае реализации ИАД представляется уместным естественный вопрос: можно ли доверять полученному результату? Однако, простейшие попытки уточнить смысл этого вопроса демонстрируют, что на самом деле все не так просто, как могло бы показаться с первого взгляда. Действительно, а что, собственно, значит «*доверять*»? Вряд ли вызовет сомнения то, что *доверие* – предмет изучения социологии (*доверяем – не доверяем, почему* именно и т.п.). *Доверенность* – качество систем и решений, изучаемое в задачах защиты данных и обеспечения информационной безопасности. *Доверительным* можно считать, в частности, определенный стиль разговора между людьми. А что с системами ИИ?

В ИИ-исследованиях и разработках представляется естественным говорить о *достаточности оснований для принятия* результатов ИАД (т.е. о «механизмах» оценивания результатов и о собственно оценках, формируемых тем или иным «механизмом»). Т.е. принимая во внимание характеристику ИАД в виде упорядоченной тройки вида  $\Sigma$ , придется согласиться, что достаточность оснований для принятия результатов ИАД связана с:

- «представительностью» (репрезентативностью) анализируемого конкретного data set'a;
- возможностями варьирования используемого ЯПЗ, и конечно же – результативностью задействованного ИАД-«инструментария».

При этом не сложно привести примеры:

- не-представительности «обучающего» data set'a в так называемом *интерполяционно-экстраполяционном* подходе (ИАД и машинное обучение, регрессионные методы САД – статистического анализа данных и др.). При работе с открытыми – пополняемыми информацией принципиально нового характера – большими данными хорошо известная в статистике проблема *репрезентативности* выборки (data set'a) приобретает абсолютно иной характер. Классические представления о *генеральной совокупности, выборках* из нее, и их *репрезентативности* в открытых мирах Big Data теряют свою корректность. (Что такое генеральная совокупность в условиях характерного для Big Data эффекта *Open?*);

- неэффективности (или же вообще неадекватности) конкретного ЯПЗ при анализе определенных типов данных. Простейший пример – попытка разделять оттенки, скажем, розового цвета на черно-белом экране монитора. (Да, чего-то можно добиться разделением уровней контрастности изображения, однако, вряд ли такое решение можно будет считать эффективным по сравнению с использованием цветного монитора). Менее тривиальный пример [Митин и др., 2020] – прогнозирование энерго-эффективности химических соединений обучением на примерах и контрпримерах (при создании новых высокоэффективных видов горючего и др.). Здесь поиск нового целевого объекта ведется на базе исследования полезных свойств нескольких десятков ранее уже хорошо изученных соединений. Использование<sup>1</sup> в качестве рабочего ЯПЗ описаний химических соединений языка графов (плоских симметризованных, пространственных, пространственных с углами и расстояниями, ...) дает одни результаты, а вот подключение релевантных квантово-химических данных (характеризующих структуры  $\pi$ -электронных облаков, экстремумы потенциальной энергии  $\pi$ -электронов при перемещении по сложным молекулярным конфигурациям анализируемых соединений и др.), представляемых системами проблемно-ориентированных расчетных квантово-химических индексов, существенным образом изменяют показатели качества результатов выполняемого ИАД.

Таким образом, представляется естественным сформулировать три вопроса, ответы на которые могли бы способствовать оценке достаточности оснований для принятия результатов ИАД в конкретной ситуации:

(В1) Что может быть содержательным (семантическим) основанием для формулирования *критерия качества data set 'a'*?

(В2) Что может быть содержательным (семантическим) основанием для формулирования *критерия качества* для выбора конкретного ЯПЗ (как для определенного класса задач, так и для конкретного data set 'a')?

(В3) Что может быть содержательным (семантическим) основанием для формулирования *критерия качества* конкретного «инструментария» ИАД (и для определенного класса задач, и для конкретного data set 'a')?

Следуя ЮИ.Журавлеву, будем рассматривать ИАД и машинное обучение (МО) как особый класс задач анализа данных, где исходный – «обучающий» – набор прецедентов *интерполируется* эмпирическими зависимостями (ЭЗ) того или иного заданного класса. Эти ЭЗ далее используют

---

<sup>1</sup> Следуя гипотезе А.М.Бутлерова, утверждавшего, что свойства химического соединения полностью определяются его химической структурой.

ся для «диагностики» вновь предлагаемых описаний прецедентов, выполняемой путем проверки *экстраполируемости* найденных зависимостей на каждый из таких новых прецедентов. При этом к машинному обучению относят интерполяционно-экстраполяционные (ИЭ) процедуры на *заданном* «обучающем» наборе прецедентов, а в ИАД рассматриваются также возможности порождения ЭЗ на *последовательностях* расширяющихся наборов *описаний* прецедентов, порождаемых пополнениями информацией о новых объектах (расширениями исходного обучающего data set'a).

Несложно убедиться, что интерполируемость исходного data set'a зависимостями заданного типа оказывается не единственным условием, требующим учета в стремлении формировать *семантически* приемлемые (корректные) ИЭ-зависимости. Пример - задачи медицинской диагностики однотипных патологий, где меры противодействия изучаемому заболеванию формируются в процессе ИАД и/или машинного обучения на описаниях пополняемых групп пациентов. Здесь идентифицируемые в процессе ИАД и/или МО факторы влияния (те, что способствуют развитию изучаемой патологии) и меры нацеленных на них терапевтических действий постоянно изменялись бы при соответствующем расширении «обучающей» коллекции описаний пациентов. (Между тем, естественно ориентироваться на «ответ» факторам *причинности* формирования соответствующей патологии, которые вряд ли могут существенно изменяться от пациента к пациенту).

Таким образом, очевидным примером дополнительного (к интерполируемости исходного data set'a) требования, которое отражало бы семантику процесса «диагностики» новых прецедентов, является анализ причинности - *наследуемости* факторов причинной обусловленности изучаемых явлений (см. задачи диагностики – медицинской, технической, ...).

Выявление факторов причинной обусловленности позволяет сформировать один из вариантов решения проблемы оценки достаточности оснований для принятия результатов ИАД, обеспечив их *объясняемость* сформированными причинно-следственными зависимостями и неформальную *интерпретируемость*. Эти причинно-следственные зависимости и позволяют формировать доверие специалистов к результатам ИАД.

Механизмы оценки достаточности оснований для принятия результатов в *замкнутых* и *открытых* мирах имеют принципиальные различия. В замкнутых мирах результативны представляемые системами уравнений «балансовые» модели универсального (по полноте описания исследуемых эффектов) характера, модели в виде полных и непротиворечивых формальных теории и др. Для открытых миров критически значим мониторинг *устойчивости* (*наследуемости*) решений. Отдельный интерес представляет анализ ситуаций, где идентифицирована *неустойчивость* – сво-

его рода смена «картины мира» (например, «фазовые переходы» и т.п.). Особую роль здесь играют конкретные «инструменты» оценивания результатов ИАД вместе с интерпретацией значений оценок, порожденных с их помощью. Пример не-сохраняемости эффектов наследуемости результатов при варьировании ЯПЗ и/или «инструментария» ИАД при анализе одного и того же data set'a будет представлен ниже (см. также работу [Забейало 2022]).

## 2. Что такое «хорошая» обучающая выборка в ИАД?

«Хорошая» обучающая выборка в ИАД: можно ли уточнить этот термин? Продолжая обсуждение, обратим внимание на два аспекта:

1) Проблема качества конкретного набора входных данных, обрабатываемых системой ИИ, далеко не часто оказывается в фокусе внимания исследователей. Обычно, к ней обращаются в ситуациях, когда возникли вопросы о приемлемости результатов. В таких случаях уместен естественный вопрос: а можно ли было вообще получить из имеющихся входных данных – обучающей выборки – заслуживающий доверия результат? И это вполне понятно, ведь знаменитую формулу *garbage in → garbage out* (мусор на входе → мусор на выходе) достаточно трудно опровергнуть!

2) В таком контексте естественно задуматься о возможностях формулирования критериев или хотя бы требований, которые позволяли бы при решении задач ИАД еще «на берегу» оценивать приемлемость исходных данных обучения. Фактически, речь – о возможности, как минимум, ответить на вопрос: *Эта выборка – «хорошая»* (приемлемая, пригодная для получения заслуживающих внимания результатов, ...), *или же она – «плохая»* (не позволяет порождать приемлемые результаты)? Дополнительно в ситуации, когда предложенная для анализа выборка оказалась «плохой» хотелось бы также получить ответ на следующий вопрос: *А что следует предпринять с этой выборкой, чтобы преобразовать ее в «хорошую»* (например, какие данные в нее дополнительно добавить, или же, наоборот, что из нее отбросить)? Некоторые аспекты этой проблемы нам и предстоит обсудить.

При переводе обсуждения оценки качества обучающей выборки в ИАД в более формализованную плоскость полезно учесть еще два аспекта:

3) Классические представления о качестве обучающей выборки, перенесенные из САД в технические приложения (наработки на отказ; генеральная совокупность, а также выборки из нее, далее - репрезентативные выборки и проблема оценки репрезентативности той или иной конкретной выборки, в т.ч. – репрезентативность, оцениваемая лишь в режиме *post factum*).

4) Открытость данных в задачах ИАД, демонстрирующая неадекватность классических представлений о репрезентативности обучающей выборки. (В т.ч. – неадекватность самого понятия *выборки*: а из чего, собственно, выбираем? Что такое здесь *генеральная совокупность*?). По-видимому, лучше говорить об обучающем наборе данных (data set) и т.п.

Сегодня уже известен ряд (не претендующих на универсальность, тем не менее, полезных и результативных) вариантов ответа на вопрос: «Что же такое «хорошая» обучающая выборка в ИАД?». Вот три примера:

(1) *Абдуктивная объясняемость* обучающего набора данных в ДСМ-методе [Финн 2010]. Здесь по результатам обработки обучающего множества прецедентов  $DS$  – примеров и контрпримеров – требуется, чтобы каждый входящий в него прецедент *объяснялся* бы по крайней мере одной из ЭЗ, порожденных на  $DS$  (*экстраполяцией* такой ЭЗ на этот пример/контрпример). Это условие называется *абдуктивной объясняемостью* исходного множества прецедентов  $DS$ . Если в  $DS$  имеются не объясненные элементы, предлагается расширить  $DS$  новыми прецедентами, сходными (в смысле лежащей в основе ДСМ-метода операции сходства описаний прецедентов) с необъясненными. Таким путем можно повысить «качество» исходного множества  $DS$  (считая, что этому повышению способствует *абдуктивная объяснимость* исходного  $DS$  порождаемыми из него ДСМ-ЭЗ).

(2) *Каузальная репрезентативность* обучающего набора данных в задачах диагностического типа [Грушо и др., 2020, Забейало 2021]. Здесь каждый элемент исходного  $DS$  представим как следствие некоторого множества ЭЗ определенного типа. Разделение примеров и контрпримеров из  $DS$  специальным логическим условием – характеристической функцией ( $XФ$ ), обращающейся в истину на каждом из примеров, и в ложь на каждом из контрпримеров в  $DS$  – означает возможность выделения в описаниях прецедентов из  $DS$  таких комбинаций факторов, причинное «влияние» которых относит прецедент к примерам, а отсутствие такового – к контрпримерам в исходном  $DS$ . *Непустота* множества  $ChF(DS)$  порождаемых на  $DS$  функций  $ChF$  представленного вида позволяет говорить о *каузальной репрезентативности* множества  $DS$ . Математическая техника  $XФ$  позволяет также описывать и эффект *наследуемости* интерполирующих исходное множество  $DS$  зависимостей из  $ChF(DS)$  при расширении  $DS$  новыми прецедентами  $\Delta DS$  и порождении нового множества  $ChF(DS \cup \Delta DS)$ . Здесь именно анализ *причинности* оказывается семантическим основанием для наследуемости ранее построенных ЭЗ на  $DS \cup \Delta DS$ . Основой для выявления соответствующих факторов причинности (и формирования причинно-следственных зависимостей) служит *эвристика* поиска *сходства*, представляющая собою конструктивный инструмент выявления скрытых в исходных данных каузальных ЭЗ. Итак, кау-

зальная репрезентативность выборки – конструктивно проверяемая характеристика. (А в [Забейайло 2021] показано, что эта проверка может быть выполнена *полиномиально* быстро).

(3) *Использование содержательных контр-эффектов* ([Забейайло и др., 2021]) для проверки фальсифицируемости ЭЗ, формируемых на обучающем множестве прецедентов (проверка «качества» исходного множества прецедентов и порождаемого из него множества причинно-следственных зависимостей попытками породить эмпирическое противоречие интерполяцией на вновь «диагностируемый прецедент одновременно двух ЭЗ, описывающих каузальную «природу» двух в реальном мире несовместимых эффектов). В работе [Забейайло и др., 2021] на эмпирических данных отделения Радиологии НМИЦ нейрохирургии им. Н.Н. Бурденко выделены два не допускающих одновременной реализации эффекта – так называемой *псевдо-прогрессии (ПсП)* некоторого вида злокачественных новообразований головного мозга человека (увеличения зоны поражения, фиксируемого объективными средствами контроля – МРТ, ПЭТ) с одной стороны, и эффекта *сокращения объема опухоли (СОО)* в периоде выполнения штатной серии сеансов лучевой терапии (сокращение зоны поражения, фиксируемое средствами объективного контроля) – с другой, демонстрируется возможность потенциальной фальсифицируемости порождаемых на одной и той же коллекции описаний пациентов *ПсП*-зависимостей средствами *СОО*-зависимостей и наоборот. В случае успеха подобной фальсификации идентифицируются артефакты ИАД и МО, обычно относимые к эффектам *переобучения*. В противном случае появляются дополнительные аргументы в пользу приемлемости «качества» накопленных на текущий момент в обучающем наборе прецедентов сведений об анализируемых патологиях и достаточности оснований для принятия результатов ИАД на текущем *DS*.

### **3. Что такое «хороший» ЯПЗ для конкретной задачи ИАД?**

Варианты ответа на этот вопрос на текущий момент разработаны существенно менее детально, чем на вопрос о качестве исходных данных. Тем не менее, можно привести соображения об *семантически связанных* Языках Представления Знаний (см., в т.ч., в Разделе 1 ситуацию с описаниями химических структур в виде тех или иных графов – плоских симметризованных, пространственных, и т.п., – дополненную использованием теоретико-множественных описаний в виде спектров излучения или поглощения). Здесь естественным образом задаваемый на семействе однородных (см. выше графовые представления химических соединений) ЯПЗ частичный порядок, характеризуемый предоставляемыми *степенями детальности* описания, может быть «транслирован» на результаты: полу-

ченные в более детальном описании заключения естественно считать предпочтительными по отношению к тем, что, сформированы более «грубым» ЯПЗ.

Полезно также сравнение результатов, полученных с помощью ЯПЗ из заранее заданного набора. Цель такого сравнения – прослеживание (где это возможно) устойчивости результатов ИАД при варьировании используемых ЯПЗ. Полезные аргументы при оценке достаточности основания для принятия результатов ИАД здесь могут дать как эффекты *наследуемости*, так и *ненаследуемости* формируемых заключений. (В последнем случае полезно выяснить: эффект *ненаследуемости* заключений на данном  $DS$  – это следствие «дефектов» ЯПЗ, или же это – отражение особого «статуса» самого множества  $DS$ , например, скрытых в нем «фазовых» переходов и т.п.).

Заслуживающие отдельного внимания перспективы для исследования открывает анализ каузальной репрезентативности анализируемого data set'a  $DS$  при варьировании ЯПЗ, используемого для «кодирования» элементов этого  $DS$ . Так, например, в работе [Забейайло 2022] описан эффект *ненаследуемости* каузальной репрезентативности изучаемого  $DS$  при смене ЯПЗ с простейшего теоретико-множественного на язык мультимножеств (где допускаются кратные вхождения одного и того же элемента в описания прецедентов анализируемого  $DS$ ).

#### 4. Что такое «хороший» инструмент (метод, алгоритм, ...) для решения конкретной задачи ИАД?

Здесь ситуация оказывается еще более запутанной. С формальной точки зрения каждая процедурная конструкция ИЭ-типа, «настраиваемая» при интерполяции на заданном «обучающем» data set'e  $DS$ , может быть описана как частично-определенная функция, значения которой зафиксированы на множестве  $DS$ , однако, за его пределами могут варьироваться, любыми допустимыми способами. Таким образом, уместны два вопроса:

(4.1) Если задано множества  $F(DS)$  допустимых частично-определенных функций  $F$  того или иного конкретного вида  $\Phi$ , корректно описывающих (интерполирующих) исходное множество прецедентов  $DS$ , как сформулировать критерий качества  $Q_{\text{INT}}$ , позволяющий выбрать наилучшую в смысле  $Q_{\text{INT}}$  для  $DS$  функцию  $F_{Q_{\text{INT}}}$  в множестве  $F(DS)$ ?

(4.2) Как в данном случае для заданного нового прецедента  $\phi$ , не входящего в исходное множество  $DS$ , сформулировать критерий качества  $Q_{\text{EXT}}$ , позволяющий выбрать наилучшую в смысле  $Q_{\text{EXT}}$  для  $DS$  функцию  $F_{Q_{\text{EXT}}}$  в продолжениях  $F(DS)$  на всю область определения функций вида  $\Phi$ ?



В контексте *каузальных оснований* для наследуемости (устойчивости) порождаемых на заданном  $DS$  интерполяций  $F$  при расширении этого текущего  $DS$  заданным дополнительным набором прецедентов  $\Delta DS$  хотелось бы также иметь возможности выделять в расширениях  $F(DS)$  на  $F(DS \cup \Delta DS)$  такие, которые представляли собою (в случае непустоты множества таких функций) до-определения частично-определенных функций из  $F(DS)$  на расширение  $\Delta DS$  области их определения. Ряд полезных фактов о реализуемости таких возможностей и характеристиках их вычислительной сложности можно найти в работах [Забежайло 2020, Забежайло 2021].

К сожалению, при попытках выделить из обширного спектра доступных на текущий момент инструментальных средств ИАД и МО наиболее подходящие для решения конкретной прикладной задачи уже на самом верхнем (общем) уровне без удовлетворительного (неоспариваемого в честном споре) ответа остается ряд принципиально важных вопросов, в т.ч.:

(?) *ПОЧЕМУ* результат применения ИАДМО-инструментария  $M$  к набору данных  $DS$  следует принять? (Например, *ПОЧЕМУ* обученная на  $DS$  искусственная нейросеть  $M$  класса  $M$  на новом – не входящем в  $DS$  прецеденте  $\phi$  – даст *правильный* результат? *ПОЧЕМУ* при заданном зашумлении  $B(\phi)$  нейросеть  $M$  вновь «распознает»  $\phi$  *правильным* образом?)

(??) *ПОЧЕМУ* из множества возможных вариантов ИАДМО-инструментария класса  $M$  к набору данных  $DS$  следует применить именно инструмент  $M$ ? (Например, *ПОЧЕМУ* это должна быть именно искусственная нейросеть  $M$  класса  $M$ ? *ПОЧЕМУ* при заданном зашумлении  $B(\phi)$  это вновь должна быть нейросеть  $M$ ?)

## Заключение

Итак: проблема оценки достаточности оснований для принятия результатов ИАД может быть рассмотрена как проблема оценки «качества» всей тройки – обучающей выборки, ЯПЗ и собственно «инструментария» ИАД. Таким образом, у оценки достаточности оснований для принятия результатов ИАД могут быть выделены три значимые составляющие «набора готовности». Как следствие, актуальны как проблема *формирования* «инструментов» оценки качества в открытых предметных областях, так и проблема *оценивания* результатов, полученных их применением. Ключевую роль здесь играют *методы* оценки вместе с собственно *результатами оценивания*, представляющие собою основания для принятия/непринятия решения о достаточности аргументов для принятия результатов ИАД.

## Список литературы

- [Грушо и др., 2021] Грушо А.А., Грушо Н.А., Забейайло М.И., Тимонина Е.Е. Поддержка решения задач диагн. типа // Информ. и ее прим. – 2021. – № 1. – С. 80-86.
- [Забейайло, 2020] Забейайло М.И. О нек. оценках сложн. выч. при прогн. свойств новых объектов средствами характ. Функций // НТИ. Сер. 2. – 2020. – № 12. – С. 1-8.
- [Забейайло, 2021] Забейайло М.И. О емкости семейств характ. функций, обеспечив. корректное решение диагн. задач // ИИ и ПР. – 2021. – № 2. – С. 44-54
- [Забейайло и др., 2021] Забейайло М.И., Трунин Ю.Ю. О значимости эмпирич. противор. в оценке надежн. рез. ИАД // НТИ. Сер. 2. – 2021. – № 6. – С. 14-20.
- [Митин и др., 2020] Митин А.В., Забейайло М.И. Инновационный подход к поиску новых веществ с заданными физ.-хим. свойствами на основе ИИ и квантово-химических расчетов // Конф, пам. ак. О.М. Белоцерковского. – М.: МФТИ, 2020.
- [Финн, 2010] Финн В.К. Индуктивные методы Д.С. Милля в системах ИИ/ИИ и принятие решений. Ч. I. – 2010. – № 3. – С. 3-21; Ч. II // Там же. – 2010. – № 4. – С.14-40.
- [Финн и др., 2021] Забейайло М.И., Михеенкова М.А., Финн В.К. ИИ: «...как вы лодку назовёте, так она и поплывёт» // 19 Нац. Конф. по ИИ (КИИ-2021, 11–16 октября 2021 г.). – Таганрог: Изд-во ЮФУ, 2021. – С. 109-119.

## О РОЛИ АРГУМЕНТАЦИИ В АНАЛИЗЕ МНЕНИЙ

М.А. Михеенкова (*m.mikheyenkova@yandex.ru*)

В.К. Финн (*v.k.finn@yandex.ru*)

Федеральный исследовательский центр  
«Информатика и управление» РАН, Москва  
РГГУ, Москва

В работе рассматриваются варианты аргументационной семантики для многозначных логик бессмысленностного и неопределённого типов. Предложенные формальные средства используются для формирования мнений на основе аргументации и анализа их соответствия объективному знанию в ситуации социологического опроса. Показана зависимость значения логической непротиворечивости мнения по отношению к такому знанию от аргументационной семантики, что свидетельствует о необходимости выбора инструментов, адекватных природе задачи.

**Ключевые слова:** аргументация, логики аргументации, многозначные логики, рациональность мнений.

### Введение

Исследования в области изучения индивидуального и/или общественного восприятия действительности (политической, социальной, культурной и т.п.) – т.е. мнений относительно различных её аспектов – в последнее время заметно сместились в сторону так называемого «анализа тональности» – *sentiment analysis* (иногда – *opinion mining*). Это связано с взрывным и неконтролируемым объёмом текстовой информации, что приводит к необходимости разработки инструментов автоматического анализа такой информации. Процедуры этого направления развиваются не только в рамках инженерно-лингвистического подхода (основанного на лингвистических словарях и правилах) [Лукашевич, 2017], но и используют современные достижения искусственного интеллекта, прежде всего – машинного обучения [Николенко и др., 2022; Wankhade et al., 2022]. Задача такого анализа – фиксация оценки (точки зрения) относительно рассматриваемой сущности и/или её аспектов без какой-либо привязки к

источникам, приводящим к становлению такой оценки (мнения). Однако для значительного ряда явлений интерес представляет не только само по себе отношение к ним, но и индивидуальные основания для его формирования. Выявление оснований обеспечивает глубину анализа рассматриваемого феномена, возможность теоретических построений и обоснованность принимаемых решений. Это относится, прежде всего, к изучению политических и электоральных предпочтений, анализу конфликтов, социальным и социо-психологическим исследованиям.

## **1. Аргументационная составляющая мнений**

Одним из основных источников получения эмпирического материала в социологических исследованиях является опрос. Современные опросные методологии проектирования опросного инструмента (опросника) включают рекомендации относительно учёта когнитивных особенностей личности. Это означает разработку способов неформального извлечения оснований представленных ответов [Садмен и др., 2003; Schwarz, 2007]. Помимо собственно понимания индивидом смысла вопросов (их семантики) основанием могут служить установки, личностный опыт, контекст, соответствие социально одобряемым взглядам и даже прагматические соображения. Осознание этих оснований респондентом с их обязательной последующей вербализацией позволяет говорить о выборе некоторой аргументации для высказанного мнения.

Аргументация в широком смысле является неотъемлемой частью как повседневной жизни и функционирования общественных институтов, так и – прежде всего – научного познания. Являясь естественным следствием ограниченности человеческого знания и, как результат, несводимости рассуждений к схемам доказательств, аргументация привлекает внимание исследователей в области искусственного интеллекта, в фокусе внимания которых находится формализация рассуждений. Это направление представлено, в частности, работами в области логических теорий аргументации [Rahwan et al., 2009; Varoni et al., 2018; Вагин, 2019; Besnard et al., 2020; Dastani et al., 2020]. Эти исследования находят множество применений, в частности, в логическом программировании и немонотонных рассуждениях [Dung, 1995; García et al., 2004], проектировании онтологий [Tempich et al., 2007]. Абстрактная теория аргументации представляет интерес для развития машинного обучения [Cocarascu et al., 2016], а в связи с исключительной ролью аргументации в правовой теории и практике занимает видное место в формальном моделировании юридического обоснования [Bench-Capon, 2020].

Существенный интерес представляет использование аргументации респондентом при формировании мнений в процессе социологического исследования<sup>1</sup>. В [Финн и др., 2020а] аргументированное отношение респондентов к теме социологического опроса  $T$  рассматривается как результат индивидуальной оценки (непосредственно в ситуации опроса) утверждений (параметров опроса) из  $P = \{p_1, \dots, p_n\}$ , раскрывающих её содержание. При этом отбор релевантной аргументации – утверждений из  $P$  – осуществляется на уровне формирования содержательной модели и является прерогативой специалиста, носителя экспертного знания. Такая аргументация оказывается составляющей творческого процесса и не поддается формализации. Здесь могут быть полезны исследования в области извлечения аргументативной составляющей из текстов на естественном языке [Stede et al., 2018].

В [Михеенкова и др., 2022] рассмотрен подход к формированию оценки утверждений  $p_1, \dots, p_n$  на основе формальной процедуры её порождения с использованием семантики логики аргументации  $A_4$  [Финн, 2020; Финн, 2021, с. 199–221]. При таком подходе аргументация позволяет выделять рациональную составляющую мнения. Эмпирическое исследование социальных явлений (поведения, установок, мнений) на основе результатов социологического опроса предполагает, по меньшей мере, решение следующих задач: выявление детерминирующих влияний (как внутренних, определяемых личностью респондента, так и внешних, характеризующих контекст явления), выделение на их основе социальных общностей с последующим прогнозированием поведения (выражения мнений) таких общностей или индивидуумов. Отдельный интерес представляет распознавание рационально обоснованных детерминаций. Адекватным инструментом для решения этих задач является ДСМ-метод автоматизированной поддержки исследований [Финн, 2021, с. 27–98, 310–436], дополненный логическими средствами для представления мнений и анализа их рациональности [Финн и др., 2020а].

Основу такого анализа составляет формальное уточнение структуры социологического опроса [Finn et al., 2011]. Закрытый социологический опрос по теме  $T$  реализуется средствами дедуктивной логики  $L$  и может быть представлен в виде  $O = \langle L, P, \Sigma, K', R \rangle$ . Здесь  $P = \{p_1, \dots, p_n\}$  – множество утверждений, характеризующих  $T$ ,  $K'$  – полученное в результате исследования стабилизированное множество ответов (при расширении множества опрашиваемых новые варианты ответов не появляются),  $R$  соответствующее  $K'$  множество респондентов.  $\Sigma = \{\psi_1, \dots, \psi_s\}$  есть множество формул  $\psi_1, \dots, \psi_s$  соответствующей логики  $L$ , представляющих в логической форме семантические зависимости между элементами  $p_1, \dots, p_n$ ,

---

<sup>1</sup> «Расскажите подробно, как Вы пришли к этому ответу...» [Садмен и др., 2003, с. 32].

заложенные исследователем на основании экспертной модели. Множество  $\Sigma$  должно быть непротиворечивым, при этом формула  $\bar{\psi} = (\psi_1 \& \dots \& \psi_n)$  не должна быть тавтологией логики  $L$  (здесь  $\&$  – соответствующая конъюнкция логики  $L$ ). Таким образом, множество  $\Sigma$  оказывается элементом формализованной эвристики. Указанные свойства  $\Sigma$  проверяются логическими средствами – методом аналитических таблиц для соответствующей логики  $L$  (ср. [Михеенкова и др., 2022])<sup>2</sup>, тогда как вопрос содержательной достаточности описания зависимостей между элементами  $P$  в  $\Sigma$  решается вне рамок формального аппарата и остаётся прерогативой специалиста. Предложенное задание множества  $\Sigma$  позволяет говорить о рациональном представлении темы с учётом объективного знания и открывает возможности для оценки рациональности восприятия темы респондентом.

## 2. Формализация аргументационной семантики

Согласно идее Д.А. Бочвара, некоторые многозначные логики являются фрагментами формализованной семантики – механизма представления знания [Финн и др., 2020]. Это означает, в частности, переход от семантики истинности к семантике понимания: истинностные значения должны иметь интерпретацию. Усилением этой идеи является представленный в [Михеенкова и др., 2022] аргументационный подход к формированию оценки утверждений из  $P$ , когда истинностное значение на основе выбранной аргументации отображает субъективное знание.

В указанной работе были рассмотрены возможные семантики истинностных значений логик бессмысленностного типа  $V_4$ ,  $E_4$  и неопределённого типа  $L'_4$ , порождённых из так называемых «логик значения»  $V_3$ ,  $E_3$  и логики  $L'_3$  [Финн и др., 2020] с использованием аргументационного подхода и сохранением интерпретации оценок высказываний, отличных от «истина» и «ложь». Так, в  $V_4$  промежуточные истинностные значения интерпретируются как сильно бессмысленные, в  $E_4$  – как слабо бессмысленные (что близко к способу употребления бессмысленных высказываний в естественном языке), в  $L'_4$  – как неопределённые, для оценки которых недостаёт информации.

Рассмотрим подход к аргументационной семантике, который можно охарактеризовать, как «гибридный». Будем исходить из того, что высказывание, имеющее одновременно аргументы «за» и «против» его принятия является неопределённым, бессмысленным же является высказывание, относительно которого аргументация отсутствует вовсе.

---

<sup>2</sup> Для непротиворечивого множества  $\Sigma$  аналитическая таблица  $\mathcal{T}_\Sigma$  (или, что то же,  $\mathcal{T}_{\bar{\psi}}$ ) не замкнута.

Итак, пусть  $A$  – множество доводов (аргументов и контраргументов), являющихся основанием принятия или непринятия респондентом некоторых утверждений (например, для  $P = \{p_1, \dots, p_n\}$  темы  $T$ ).  $P$  – множество всех пропозициональных переменных. Для каждой  $p_i \in P$  определим две функции  $g^+$  и  $g^-$ , задающие множество аргументов «за» и множество аргументов «против», соответственно:

$$g^+ : P \rightarrow 2^A, g^+(p_i) \subseteq A, i = 1, \dots, n,$$

$$g^- : P \rightarrow 2^A, g^-(p_i) \subseteq A, i = 1, \dots, n.$$

Рассмотрим множество истинностных значений  $V = \{1, -1, u, \tau\}$ , интерпретируемых как «истина», «ложь», «неопределённость», «бессмыслица», соответственно.

Зададим атомарную оценку для принятия утверждений  $p_i \in P$  ( $i = 1, \dots, n$ ), выбирая аргументы из аргументационной базы  $A$ :

$$v[p_i] = 1 \Leftrightarrow g^+(p_i) \neq \emptyset \ \& \ g^-(p_i) = \emptyset;$$

$$v[p_i] = -1 \Leftrightarrow g^+(p_i) = \emptyset \ \& \ g^-(p_i) \neq \emptyset;$$

$$v[p_i] = u \Leftrightarrow g^+(p_i) \neq \emptyset \ \& \ g^-(p_i) \neq \emptyset;$$

$$v[p_i] = \tau \Leftrightarrow g^+(p_i) = \emptyset \ \& \ g^-(p_i) = \emptyset.$$

Заметим, что аргументационные функции каждого респондента индивидуальны, т.е. для множества респондентов  $R = \{r_1, \dots, r_m\}$  задаются аргументационные функции  $g_j^\sigma(p_i), j = 1, \dots, m, i = 1, \dots, n, \sigma \in \{+, -\}$ , причём для рационально мыслящего субъекта  $g_j^+(p_i) \cap g_j^-(p_i) = \emptyset$ .

На основании выбранных аргументационных функций респондент формирует оценку  $v$  для утверждений  $p_i$  ( $i = 1, \dots, n$ ),  $v \in \{1, -1, u, \tau\}$ . Составом мнения  $j$ -го респондента в опросе  $O$  будем называть множество

$$\text{атомов } \{J_{v_1^{(j)}} p_1, \dots, J_{v_n^{(j)}} p_n\}, \text{ где } J_{v_i^{(j)}} p_i = \begin{cases} t, & \text{если } v^{(j)}[p_i] = v_i^{(j)} \\ f, & \text{если } v^{(j)}[p_i] \neq v_i^{(j)} \end{cases},$$

$v_i^{(j)}$  – соответствующие оценки  $j$ -го респондента для утверждений  $p_i$  ( $i = 1, \dots, n$ )  $v_i^{(j)} \in \{1, -1, u, \tau\}, j = 1, \dots, k$ . Здесь  $J_v$  – введённые Д.А. Бочваром логические связки, которым соответствуют характеристические функции, распознающие истинностные значения  $v, v[p]$  – функция оценки,  $t$  и  $f$  – (внешние) истинностные значения двузначной логики «истина» и «ложь», соответственно. Мнением  $\varphi_j$   $j$ -го респондента будем называть максимальную конъюнкцию соответствующей логики  $L$  опроса  $O$ ,  $\varphi_j = J_{v_1^{(j)}} p_1 \& \dots \& J_{v_n^{(j)}} p_n$  ( $\equiv$  – графическое равенство), определяемую аналогично максимальной конъюнкции двузначной логики. Множество всех возможных ответов в опросе  $O$  по теме  $T$  с  $P = \{p_1, \dots, p_n\}$  обозначим  $K, |K| = 4^n, K = \{\varphi_j \mid \varphi_j = J_{v_1^{(j)}} p_1 \& \dots \& J_{v_n^{(j)}} p_n, v^{(j)}[p_i] = v_i^{(j)}, v_i^{(j)} \in \{1, -1, u, \tau\}, i = 1, \dots, n, j = 1, \dots, 4^n\}$ . В опросе  $O$  реализуются лишь некоторые из них, стабилизированное множество ответов  $K' \subseteq K, |K'| \leq m$ .

Можно установить, соотносится ли мнение  $j$ -го респондента с представленными в  $\Sigma$  зависимостями между утверждениями  $p_1, \dots, p_n$ , которые отражают рациональное знание эксперта о предметной области. Это достигается проверкой непротиворечивости множества  $\{\psi_1, \dots, \psi_s, \varphi_j\}$ , что позволяет уточнить, насколько рациональным является выбор  $j$ -м респондентом аргументационных функций  $g_j^\sigma(p_i)$  ( $j = 1, \dots, m, i = 1, \dots, n, \sigma \in \{+, -\}$ ) для элементов  $P = \{p_1, \dots, p_n\}$  и служит маркером рациональности респондента.

Пусть  $consis(\Sigma)$  – метаязык непротиворечивости множества формул ( $\Sigma$ ), которая, как уже говорилось выше, проверяется методом аналитических таблиц соответствующей логики  $L$  опроса  $O$ . Рассмотрим полученное в результате опроса  $O$  множество ответов  $K' = \{C_1, \dots, C_k\}, k \leq m, K' \subseteq K, C_j$  – максимальная конъюнкция, представляющая соответствующее мнение ( $j = 1, \dots, k$ ). Построим аналитическую таблицу  $\mathcal{T}_{\Sigma \cup \{C_j\}}$  с корнем  $(\Sigma \cup \{C_j\})$ . Если она не замкнута, имеет место  $consis(\Sigma \cup \{C_j\})$ , и респондентов, имеющих такое мнение (их может быть несколько), можно считать рациональными. Для исследователя может представлять интерес отбор не противоречащих  $\Sigma$  мнений  $K^* = \{X \mid consis(\Sigma \cup \{X\}) \& (X \in K')\}, K^* \subseteq K'$  и выделение соответствующего сообщества рациональных респондентов с последующим анализом детерминаций рациональности мнений и выбора соответствующих аргументационных функций.

### 3. Исследовательские стратегии анализа рациональности

Для предложенной интерпретации аргументационно заданных бессмысленности и неопределённости можно рассмотреть логики  $LB_4, BL_4, LE_4, EL_4$  с множеством определённых выше внутренних истинностных значений  $V = \{1, -1, u, \tau\}$  (для представления фактов), внешних истинностных значений  $\{t, f\}$  (для доказательств утверждений о фактах) и расширенной импликацией Бочвара  $\rightarrow$ . Отрицание  $\sim$  для всех логик определяется одинаково (ср. с  $\sim$  для  $B_4, E_4, L_4$  в [Михеенкова и др., 2022]):

$p$	$\sim p$	$\rightarrow$	$f$	$-1$	$\tau$	$u$	$1$	$t$
$f$	$t$	$f$	$t$	$t$	$t$	$t$	$t$	$t$
$-1$	$1$	$-1$	$t$	$t$	$t$	$t$	$t$	$t$
$\tau$	$u$	$\tau$	$t$	$t$	$t$	$t$	$t$	$t$
$u$	$\tau$	$u$	$t$	$t$	$t$	$t$	$t$	$t$
$1$	$-1$	$1$	$f$	$f$	$f$	$f$	$t$	$t$
$t$	$f$	$t$	$f$	$f$	$f$	$f$	$f$	$t$



Логические связки – конъюнкция  $\&$  и дизъюнкция  $\vee$  – определяются для каждой логики соответствующими истинностными таблицами: для истинностных значений  $1, -1, \tau$  учитываются соответствующие истинностные таблицы логики  $B_3$  и/или  $E_3$ , для  $1, -1, u$  – логики  $L_3$ , для  $t$  и  $f$  – двузначной логики  $L_2$ .

Полное описание логического аппарата находится за рамками нашего изложения. Приведём лишь примеры некоторых истинностных таблиц.

В логике  $LB_4$  для  $\vee$  и  $\&$  учитываются истинностные значения  $\tau, -1, 1$  ( $B_3$ ), и порядок  $-1 < u < 1$  ( $L_3$ ),  $t$  и  $f$  ( $L_2$ ). Здесь  $\tau \vee z = \tau, \tau \& z = \tau, f \& z = f, t \& z = z, t \vee z = t, f \vee z = z, z \in \{1, -1, u, \tau\}$ . Соответственно, истинностные таблицы для  $\vee$  и  $\&$  имеют вид

$\vee$	$f$	$-1$	$\tau$	$u$	$1$	$t$
$f$	$f$	$-1$	$\tau$	$u$	$1$	$t$
$-1$	$-1$	$-1$	$\tau$	$u$	$1$	$t$
$\tau$	$\tau$	$\tau$	$\tau$	$\tau$	$\tau$	$t$
$u$	$u$	$u$	$\tau$	$u$	$1$	$t$
$1$	$1$	$1$	$\tau$	$1$	$1$	$t$
$t$	$t$	$t$	$t$	$t$	$t$	$t$

$\&$	$f$	$-1$	$\tau$	$u$	$1$	$t$
$f$	$f$	$f$	$f$	$f$	$f$	$f$
$-1$	$f$	$-1$	$\tau$	$u$	$-1$	$-1$
$\tau$	$f$	$\tau$	$\tau$	$\tau$	$\tau$	$\tau$
$u$	$f$	$u$	$\tau$	$u$	$u$	$u$
$1$	$f$	$-1$	$\tau$	$u$	$1$	$1$
$t$	$f$	$-1$	$\tau$	$u$	$1$	$t$

В логике  $BL_4$  учитывается порядок истинностных значений  $-1 < u < 1$  ( $L_3$ ), значения  $\tau, -1, 1$  ( $B_3$ ),  $t$  и  $f$  ( $L_2$ ). Здесь  $\tau \vee u = u, \tau \& u = \tau$ . Истинностные таблицы для  $\vee$  и  $\&$  имеют вид

$\vee$	$f$	$-1$	$\tau$	$u$	$1$	$t$
$f$	$f$	$-1$	$\tau$	$u$	$1$	$t$
$-1$	$-1$	$-1$	$\tau$	$u$	$1$	$t$
$\tau$	$\tau$	$\tau$	$\tau$	$u$	$1$	$t$
$u$	$u$	$u$	$u$	$u$	$1$	$t$
$1$	$1$	$1$	$1$	$1$	$1$	$t$
$t$	$t$	$t$	$t$	$t$	$t$	$t$

$\&$	$f$	$-1$	$\tau$	$u$	$1$	$t$
$f$	$f$	$f$	$f$	$f$	$f$	$f$
$-1$	$f$	$-1$	$\tau$	$-1$	$-1$	$-1$
$\tau$	$f$	$\tau$	$\tau$	$\tau$	$\tau$	$\tau$
$u$	$f$	$-1$	$\tau$	$u$	$u$	$u$
$1$	$f$	$-1$	$\tau$	$u$	$1$	$1$
$t$	$f$	$-1$	$\tau$	$u$	$1$	$t$

Аналогично могут быть построены истинностные таблицы для дизъюнкции  $\vee$  и конъюнкции  $\&$  для логик  $LE_4, EL_4$ , при этом на истинностных значениях  $\tau, -1, 1$  учитывается порядок  $\tau < -1 < 1$ .

Для реализации опроса  $O = \langle L, P, \Sigma, K', R \rangle$  и анализа рациональности мнений требуется расширение аппарата этих логик добавлением одноместных логических связок  $J_{-1}, J_1, J_u, J_\tau$  (как металогических операций), при этом  $J_\tau$  и  $J_f$  выразимы посредством импликации  $\rightarrow$  и отрицания  $\sim$ :  $J_\tau p \Leftrightarrow t \rightarrow p, J_f p \Leftrightarrow t \rightarrow \sim p$ . Такие расширенные логики могут быть формализованы методом аналитических таблиц, что позволит анализировать непротиво-

речивость множества  $\Sigma$  и устанавливать рациональность мнений, т.е. формировать множество  $K^* = \{X \mid \text{consis}(\Sigma \cup \{X\}) \& (X \in K')\}$ . В [Михеенкова и др., 2022] было показано, что для опроса  $O = \langle L, P, \Sigma, K', R \rangle$  по теме  $T$  с характеризующим её содержание множеством  $P = \{p_1, \dots, p_n\}$ , участниками опроса из множества  $R$  и стабилизированным множеством ответов  $K'$  множество  $K^*$  зависит от логики  $L$ , формализующей опрос. Представленные в  $\Sigma$  связи между утверждениями  $p_1, \dots, p_n$  семантически эквивалентны, однако правила вывода в методе аналитических таблиц связаны с определением логических связей в соответствующей логике. Соответственно, результаты анализа непротиворечивости мнений отличаются, в результате чего порождаются различные множества  $K^*$ , объединяющие различные множества рациональных респондентов. Это говорит о необходимости обоснованного выбора средств формализации для адекватного отображения семантики предметной области в целом и рассматриваемой задачи в частности.

В представленной схеме опроса формирование мнений из множества  $K' = \{C_1, \dots, C_k\}$  опирается на выбор  $j$ -м респондентом аргументационных функций  $g_j^\sigma(p_i)$  ( $j = 1, \dots, m$ ,  $i = 1, \dots, n$ ,  $\sigma \in \{+, -\}$ ) для элементов  $P = \{p_1, \dots, p_n\}$ . Этот выбор – согласуется с имеющимися у респондента знаниями, и можно говорить о том, что утверждения  $p_1, \dots, p_n$  оцениваются посредством когерентной теории истины. Если же оценки утверждений  $p_1, \dots, p_n$  в опросе по той же теме  $T$ , с теми же участниками из  $R$ , также формализуем средствами логики  $L$  для представления  $\Sigma$  и анализа рациональности, есть результат непосредственной реакции респондентов в ситуации опроса, в этом случае можно говорить о взаимодействии корреспондентной и когерентной теории истины (см. [Финн, 2021, с. 27–98]). Последняя находит отражение на этапе формирования множеств  $P$  и  $\Sigma$ . Естественным следствием такой стратегии опроса является то, что множества  $K'$  и  $K^*$  предсказуемо окажутся иными, нежели в случае аргументационного формирования мнений.

В описанных логиках аргументационная семантика представлена на уровне формирования оценок атомарных высказываний  $p_1, \dots, p_n$ . Аргументационный подход может быть углублён, если распространить аргументационный подход также на «молекулярный» уровень. Иными словами, аргументационные функции для сложных высказываний формируются из аргументационных функций атомарных высказываний. Пусть, к примеру,  $v[p_i] = 1$ , т.е.  $g^+(p_i) \neq \emptyset$  и  $g^-(p_i) = \emptyset$ ;  $v[p_j] = u$ , т.е.  $g^+(p_j) \neq \emptyset$  и  $g^-(p_j) \neq \emptyset$ . Тогда  $g^\sigma(p_i \vee p_j)$  есть  $g^\sigma(p_i) \cup g^\sigma(p_j)$  ( $\sigma \in \{+, -\}$ ) и, соответственно,  $v[p_i \vee p_j] = u$ ; и т.п. Назовём логику с такой семантикой, внутренними ис-

тинностными  $\{1, -1, u, \tau\}$  и внешними истинностными значениями  $\{t, f\}$   $\tilde{A}_4$ . Истинностные таблицы для отрицания  $\sim$  и импликации  $\rightarrow$  совпадают с приведёнными выше таблицами для логик  $LB_4, BL_4, LE_4, EL_4$ . Истинностные таблицы для  $\vee$  и  $\&$  задаются следующим образом:

$\vee$	$f$	$-1$	$\tau$	$u$	$1$	$t$	$\&$	$f$	$-1$	$\tau$	$u$	$1$	$t$
$f$	$f$	$-1$	$\tau$	$u$	$1$	$t$	$f$	$f$	$f$	$f$	$f$	$f$	$f$
$-1$	$-1$	$-1$	$\tau$	$u$	$1$	$t$	$-1$	$f$	$-1$	$f$	$f$	$f$	$-1$
$\tau$	$\tau$	$\tau$	$\tau$	$u$	$1$	$t$	$\tau$	$f$	$f$	$\tau$	$f$	$f$	$\tau$
$u$	$u$	$u$	$u$	$u$	$1$	$t$	$u$	$f$	$f$	$f$	$u$	$f$	$u$
$1$	$1$	$1$	$1$	$1$	$1$	$t$	$1$	$f$	$f$	$f$	$f$	$1$	$1$
$t$	$t$	$t$	$t$	$t$	$t$	$t$	$t$	$f$	$-1$	$\tau$	$u$	$1$	$t$

Из таблиц видно, что  $\sim$  является внутренней связкой,  $\rightarrow$  – внешней,  $\&$  и  $\vee$  – смешанными. Для формализации опроса и анализа рациональности мнений логика  $\tilde{A}_4$  также должна быть расширена добавлением одноместных логических связок  $J_{-1}, J_1, J_u, J_t; J_t$ , и  $J_f$  по-прежнему выразимы. Такое расширение позволяет формализовать логику  $\tilde{A}_4$  методом аналитических таблиц, сформировать для опроса  $O = \langle \tilde{A}_4, P, \Sigma, K', R \rangle$  стабилизированное множество ответов  $K'$  и построить множество  $K^*$ , также отличные от множеств, полученных описанными выше стратегиями опроса.

Можно предложить ещё более изощрённую стратегию опроса в соответствии с предложенной в [Финн, 2021, с. 199–221] идеей многоуровневой аргументации. Аргументационные функции в многоуровневых аргументационных системах представлены отображениями  $g^\sigma: P \rightarrow 2^P$  ( $\sigma \in \{+, -\}$ ), где  $P = P^a \cup P^b \cup P'$ . Здесь  $P'$  – множество аргументирующих (базисных) высказываний,  $P^a$  – аргументируемых высказываний,  $P^b$  – аргументируемых и аргументирующих высказываний. Используя для опроса с таким множеством  $P$  логику  $\tilde{A}_4$ , т.е. аргументацию как для атомарных, так и для сложных высказываний, мы можем получить чрезвычайно сильный инструмент для анализа рациональной составляющей мнения.

## Заключение

Заметим, что даже достаточно простая аргументационная схема опроса, когда атомарные высказывания из  $P = \{p_1, \dots, p_n\}$  служат аргументами для формирования отношения к теме опроса  $T$ , в реальном исследовании электоральных предпочтений позволила обнаружить нетривиальный эффект некоммутативности результатов анализа рациональности в зависимости от стратегии проведения опроса [Финн, 2021, с. 273–289]. Оказалось, что значения степени рациональности, определяемой как  $\delta(K^*, K') = \frac{|K^*|}{|K'|}$ , различа-

ются, если сначала проводится опрос по теме, потом по  $P$ , и наоборот. Это свидетельствует о содержательности и адекватности использования аргументационной семантики в опросных технологиях.

Использование различных логик семантически оправдано изучением различных аспектов человеческой рациональности. Если основной фокус исследования направлен на изучение собственно рациональности (в обыденном понимании – логичности), разумно использовать логики  $LB_4$  или  $LE_4$ , если на анализ информированности –  $BL_4$  или  $EL_4$ . Использование логики  $\tilde{A}_4$  позволяет глубоко изучить все стороны рациональности, однако и трудоёмкость проведения опроса с её использованием достаточно велика. Эпистемологическое значение полученных результатов свидетельствует о глубине идей И. Канта об относительности знания, которое оказывается зависимым от познавательного процесса, в частности – от используемого инструмента.

Заметим также, что концепция эмпирической качественной причинности, разработанная в рамках ДСМ-метода автоматизированной поддержки исследований, представляет собой эффективную реализацию расширенного понятия рациональности средствами ИС. Образцом рациональности в рамках этого расширенного понятия является исследование, принятие результатов которого опирается на формализованную эвристику, аргументацию и преобразование неясных идей в точно определённые понятия. В предложенном подходе к анализу мнений реализуются все эти три составляющие: представление формальной структуры опроса является точным определением самого понятия опроса, элементы формализованной эвристики прослеживаются в  $\Sigma$ , формирование мнений и его анализ опирается на аргументационную семантику.

### Список литературы

- [Бочвар и др., 2020] Бочвар Д.А., Финн В.К. Некоторые дополнения к статьям о многозначных логиках // В сб.: Многозначные логики и их применения. Т. 1: Логические исчисления, алгебры и функциональные свойства / под ред. В.К. Финна). – М.: URSS, 2020. – С. 111-171.
- [Вагин, 2019] Элементы теории аргументации и её роль в интеллектуальном анализе данных // В кн.: Вагин В.Н. Знания и убеждения в интеллектуальном анализе данных. – М.: Физматлит, 2019. – С. 260-377.
- [Лукашевич, 2017] Лукашевич Н.В. Автоматические методы анализа тональности // В кн.: Автоматическая обработка текстов на естественном языке и анализ данных: учеб. пособие. – М.: Изд-во НИУ ВШЭ, 2017. – С. 127-194.
- [Михеенкова и др., 2022] Михеенкова М.А., Финн В.К. Об аргументационном подходе к анализу мнений // Сборник научных трудов XI Международной научно-технической конференции «Интегрированные модели и мягкие вычисления в искусственном интеллекте», г. Коломна, 2022В 2-х т. Т. 2..
- [Николенко и др., 2022] Как научить компьютер читать // В кн.: Николенко С.И., Кадуриин А.А., Архангельская Е.О. Глубокое обучение. Погружение в мир нейронных сетей. – СПб.: Питер, 2022. – С. 278-329.

- [Садмен и др., 2003] Как люди отвечают на вопросы: применение когнитивного анализа в массовых обследованиях. – М.: Институт Фонда «Общественное мнение», 2003.
- [Финн, 2020] Финн В.К. Стандартные и нестандартные логики аргументации // В сб.: Многозначные логики и их применения. Т. 2: Логики в системах искусственного интеллекта / под ред. В.К. Финна. – М.: URSS, 2020. – С. 59-91.
- [Финн, 2021] Финн В.К. Интеллект, информационное общество, гуманитарное знание и образование. – М.: ЛЕНАНД, 2021.
- [Финн и др., 2020] Финн В.К., Аншаков О.М., Григолия Р.Ш., Забежайло М.И. Многозначные логики как фрагменты формализованной семантики // В сб.: Многозначные логики и их применения. Т. 1: Логические исчисления, алгебры и функциональные свойства / под ред. В.К. Финна. – М.: URSS, 2020. – С. 323-356.
- [Финн и др., 2020а] Финн В.К., Михеенкова М.А. О логических средствах концептуализации анализа мнений // В сб.: Многозначные логики и их применения, т. 2: Логики в системах искусственного интеллекта / под ред. В.К. Финна. – М.: URSS, 2020. – С. 152-199.
- [Baroni et al., 2018] P. Baroni, Gabbay D., Giacomin M., van der Torre L. (eds). Handbook of Formal Argumentation. – London: College Publications, 2018.
- [Bench-Capon, 2020] Bench-Capon T.J.M. Before and after Dung: Argumentation in AI and Law // Argument & Computation. – 2020. – Vol. 11 (1-2).
- [Besnard et al., 2020] Besnard P., Cayrol C., Lagasque-Schiex M.-C. Logical theories and abstract argumentation: A survey of existing works // Argument & Computation. – 2020. – Vol. 11 (1-2).
- [Cocarascu et al., 2016] Cocarascu O., Toni F. Argumentation for machine learning: A survey // Proc. of the 6th International Conference on Computational Models of Argument, Potsdam University, Germany. 2016.
- [Dung, 1995] Dung P.M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games // Artificial Intelligence. – 1995. – Vol. 77.
- [Dastani et al., 2020] Dastani M., Dong H., van der Torre L. (eds). Logic and Argumentation // Lecture Notes in Artificial Intelligence. – 2020. – Vol. 12061.
- [Finn et al., 2011] Finn V. K., Mikheyenkova M. A. Plausible Reasoning for the Problems of Cognitive Sociology // Logic and Logical Philosophy. – 2011. – Vol. 20.
- [Garcia et al., 2004] García A.J. and Simari G.R. Defeasible logic programming: An argumentative approach // Theory and Practice of Logic Programming. – 2004. – 4 (2).
- [Rahwan et al., 2009] Rahwan I., and Simari G.R. (Eds.) Argumentation in artificial intelligence. – Dordrecht, Heidelberg, London, New York: Springer, 2009.
- [Schwarz, 2007] Schwarz N. Cognitive aspects of survey methodology // Applied Cognitive Psychology. – 2007. – Vol. 21 (2).
- [Stede et al., 2018] Stede M. and Schneider J. Argumentation Mining. – Morgan & Claypool Publishers, 2018.
- [Tempich et al., 2007] Tempich C., Simperl E., Luczak M., Studer R. and Pinto H. Argumentation-based ontology engineering // IEEE Intelligent Systems. – 2007. – Vol. 22 (6).
- [Wankhade et al., 2022] Wankhade M., Rao A.C.S., Kulkarni C. A survey on sentiment analysis methods, applications, and challenges // Artificial Intelligence Review. – Springer Nature, 2022.

## ГИБРИДИЗАЦИЯ ОНТОЛОГИЙ И НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ ДЕТЕКТИРОВАНИЯ АНОМАЛИЙ ВРЕМЕННЫХ РЯДОВ\*

В.С. Мошкин (*postforvadim@ya.ru*)

Д.С. Курило (*dimokkurilo@gmail.com*)

И.А. Андреев (*ares-ilya@ya.ru*)

Ульяновский государственный технический университет,  
Ульяновск

В статье описаны результаты разработки алгоритма обнаружения аномалий временных рядов с учетом особенностей предметной области. Алгоритм предполагает нахождение прогноза временных рядов с использованием LSTM-сетей, обнаружение аномалий по полученному прогнозу, фильтрацию найденных аномалий в соответствии с возможными отклонениями значений временного ряда от тренда, отраженными в онтологии, и логический вывод результатов поиска с использованием набора SWRL-правил. Эффективность предложенного подхода подтверждена рядом экспериментов, проводимом на бенчмарке данных по работе нефтяных вышек.

**Ключевые слова:** временные ряды, онтология, нейронная сеть, детектирование аномалий.

### Введение

Поиск и обнаружение аномалий временных рядов – это область интеллектуального анализа данных, позволяющая находить точки в  $n$ -мерном пространстве, выбивающиеся из общей тенденции поведения временных рядов. Эти числовые значения могут относиться к самым различным предметным областям и могут говорить о различных проблемах, сбоях или данных, на которые стоит обратить внимание [Lindigrin, 2021] [Hasanov, 2018].

---

\* Работа выполнена при финансовой поддержке Минобрнауки России в рамках проекта №075-00233-20-05 от 03.11.2020 «Исследование интеллектуального предиктивного мульти-модального анализа больших данных и извлечения знаний из различных источников».

Для выполнения задачи детектирования аномалий в настоящее время используются различные технологии, которые можно объединить в три основных класса:

1. Методы распознавания аномалии без учителя.

Данный набор алгоритмов используется в тех случаях, если по условиям задачи неизвестен характер временных рядов и неопределенно наличие аномалий. В этом случае алгоритмы базируются на предположении, что аномальные значения встречаются реже нормальных. Данные подходы требуют анализа данных в реальном времени [Busseti, 2012] [Chandola, 2007] [Ahmad, 2016].

2. Методы распознавания аномалии частично с учителем.

Данный набор алгоритмов предполагает наличие обучающего множества, в котором определены только нормальные объекты (не аномальные). В данном случае алгоритмы предполагают, что объекты, которые не относятся к классу нормальных, являются аномальными [Das, 2009] [Krizhevsky, 2010] [Nikolov, 2010].

3. Методы распознавания аномалий с учителем.

Данный набор алгоритмов предполагает наличие полностью размеченного обучающего множества на класс нормальных объектов и аномальных объектов. В большинстве случаев предполагается, что данные не меняют свои статистические характеристики, так как возникает необходимость изменять классификатор [Sak, 2014] [Liu, 2016].

Главной сложностью применения этих подходов является необходимость формирования данных для обучения, особенно актуально это при обучении классов аномальных данных

Главным недостатком всех перечисленных групп алгоритмов является абстрагирование от особенностей предметной области, так как анализируются только численные значения. В свою очередь аномалия, применимая к одной предметной области, может являться нормальным объектом для иной области.

В связи с этим, актуальной является задача гибридизация эффективных алгоритмов детектирования аномалий временных рядов с семантическими (онтологическими) алгоритмами, позволяющими учесть характерные особенности анализируемой предметной области.

## **1. Гибридный подход детектирования аномалий**

В рамках проекта был разработан алгоритм поиска аномалий во временных рядах с учетом особенностей предметной области. Базовый алгоритм анализа состоит из следующих этапов, показанных на рис. 1.



Рис. 1. Блок-схема алгоритма детектирования аномалий

В качестве механизма, обеспечивающего анализ и прогнозирование временных рядов (ВР), используется нейронная сеть оригинальной архитектуры [Zuev, 2021]. Первыми шагами алгоритма является предобработка данных, обучение и использование этой сети для получения прогнозных значений. Проверяемые на аномальность значения сравниваются с прогнозными. При выявлении отклонения от прогноза данные используются далее. База знаний, обеспечивающая семантическую фильтрацию



результатов анализа ВР и логический вывод рекомендаций, представляет собой OWL-онтологию с интегрированным набором SWRL-правил [Jean-Baptiste, 2021]. Значения полученные ранее загружаются в эту онтологию в виде индивида со связями с датой замера и классом замера.

Основными развилками в работе алгоритма является проверка значения на аномальность на этапе анализа данных нейронной сетью и на этапе работы с онтологией при прохождении предыдущего условия. Вне зависимости от исхода проверок все данные попадают в логический вывод результатов, завершающий работу системы [Wen, 2019].

Нейронная сеть включает в себя следующие слои:

- ConvLSTM1D – Аналогично слою LSTM, но входные преобразования и рекуррентные преобразования являются сверточными [Zhong, 2020];
- Conv1DTranspose – транспонированный свёрточный слой [Malhotra, 2015],
- Dropout – слой сброса для предотвращения переобучения.

Общая структура сети показана на рис. 2.

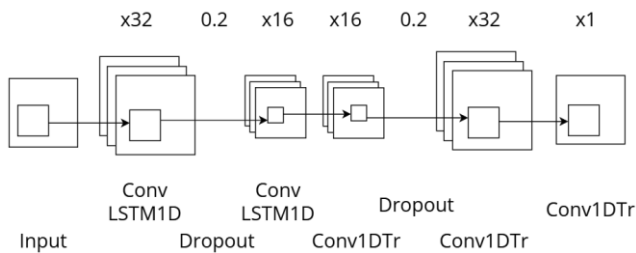


Рис. 2. Структура нейронной сети

Для использования онтологических методов в поиске аномалий был создан набор продукционных правил, определяющих нормальное поведение экземпляра временного ряда профиля. Объект, не соответствующий правилам, является аномальным. При использовании системы на реальных данных, правила должны быть предзаполнены экспертом предметной области.

Набор правил, которые обеспечивают логический вывод результатов анализа, можно разделить на две группы:

1. Группа продукционных правил, консеквентом (следствием) которых являются результат детектирования аномалии [Yarushkina, 2019]. Модель подобного правила представлена в виде выражения:

$$p_1(x, d) \wedge p_2(d, s) \wedge \dots \wedge p_n(x, c_1) \wedge g(c_1, c_2) \wedge \dots \rightarrow p_n(x, h).$$

2. Группа правил, следствием которых является логическое присвоение одной из метрики анализа ВР конкретного значения. Далее это значение выступает в роли входных данных других продукционных правил в процессе генерации логического вывода [Yarushkina, 2016]. Модель подобного SWRL-правила представлена в виде выражения:

$$p_1(x, d) \wedge p_2(d, s) \wedge \dots \wedge p_i(x, c_1) \wedge g(c_1, c_2) \wedge \dots \rightarrow p_n(x, c_n) \wedge g_n(c_n, c_{n+1}),$$

где  $x, h \in H, p_1, p_2, \dots, p_i \dots p_n \in P, g \dots g_n \in G, d \in D, c_1, c_2 \dots c_n, c_{n+1} \in C, H, P, G, D, C$  – классы замеров ВР и их свойства описанные в онтологии

Схематично описанные виды правил представлены на рис. 3.

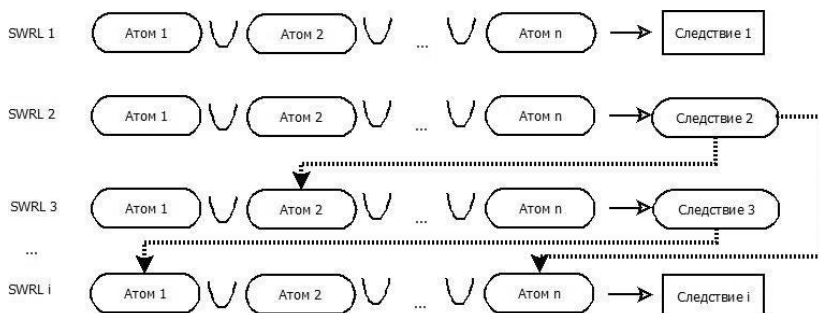


Рис. 3. Группы SWRL-правил

## 2. Подготовка данных для проведения экспериментов

Данные для проверки эффективности разработанного подхода взяты из открытых интернет-источников и содержат данные по производительности буровых вышек. Исходные данные содержат измерения для 5 различных скважин и в общей сложности более 3000 значений ВР нескольких переменных.

Для экспериментов использовались следующие данные: GR – измерение гамма-излучения, Delta PИИ – индекс пористости в петрофизике, PLD\_log10 - измерение удельного сопротивления. Пример исходных данных показан на рис. 4.

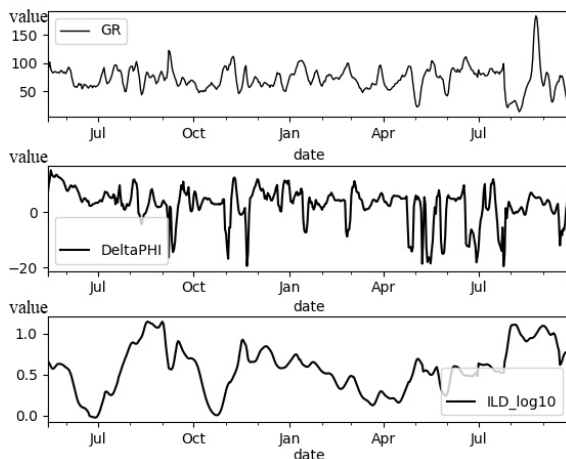


Рис. 4. Пример исходных данных

В исходных данных даты замеров отсутствуют, но происходят последовательно, через равные промежутки времени. Принято, что дата представляется, начиная с текущего дня до окончания строк таблицы с шагом один день. Даты в индексах позволяют быстрее строить графики и получить из других модулей сразу дату аномалии.

Данные разбиваются на несколько независимых друг от друга выборок по имени вышки, на которой происходили замеры. Это позволяет иметь несколько массивов данных, подходящих для последующих экспериментов. Полученные данные при необходимости возможно отобразить на графике. Готовые данные передаются в модуль нейронной сети.

### 3. Структура онтологии

Для создания онтологии использовалась ПО Protégé версии 5.5.0. Структура онтологии включает в себя несколько классов, показанных на рис. 5.

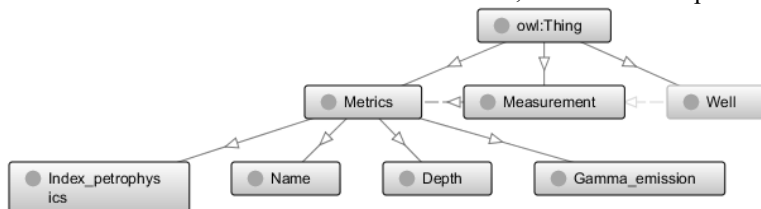


Рис. 5. Структура онтологии

В модуле онтологий содержатся:

- классы измерений;
- свойства измерений;
- правила.

Основные классы онтологии:

- Класс *Well* – объект реального мира, привязывающий измерения к предметной области.
- *Measurment* – класс показывающий, что значение является замером и частью временного ряда измерений.
- Класс *Metrics* показывает связь между замером и всеми показателями предметной области.

Правила содержат атомы класса и несколько необходимых свойств: аномального значения и дополнительного параметра, по которому подтверждается корректность суждения об аномальности.

SWRL-правила обеспечивают проверку аномалий на корректность в зависимости от характера предметной области. Пример такого правила представлен на рис. 6.

```
hasGR(?x, ?gr) ^ Well(?x) ^ hasDeltaPHI(?x, ?deltaphi) ^  
swrlb:lessThan(?deltaphi, 2.0) -> hasAnomaly(?x, true)
```

*Рис. 6. Пример SWRL-правила*

Представленное SWRL-правило использует значение временного ряда и делает вывод по параметру, не используемому в нейронной сети.

Задача правила состоит в проверке достижения параметром константы. Если значение не достигнуто, то замер может считаться аномальным, иначе значение корректно. Проверка осуществляется с помощью встроенных функций библиотеки *sqrlb*. При истинности всех предикатов индивиду в свойство аномальности ставится *true* [Moshkin, 2022].

#### **4. Проведение и результаты экспериментов**

Загрузка и выгрузка значений временного ряда в онтологию в системе происходит с помощью библиотеки *owlready2*. Она позволяет получить из OWL-файла структуру и объекты онтологии и взаимодействовать с ними на языке *python*.

Первым шагом работы является очистка всех уже существующих индивидов в онтологии. Каждый запуск загружает все найденные нейронной сетью аномалии в онтологию, а после работы правил выгружает все имеющиеся в онтологии значения.

Для загрузки индивидов в онтологию используется класс и свойства, уже описанные в структуре онтологии. После загрузки всех аномалий, происходит пересохранение онтологии и запуск правил. После обработки из файла можно получить все индивиды. В цикле проверяется флаг аномальности и отбираются все значения, принятые онтологией за действительно аномальные (рис. 7).

SWRL-правила проверяют полноту различных метрик, а также необходимое значение петрофизического индекса для достижения порогового значения 0,5. Они используют значение временного ряда и делают вывод на основе параметра, не используемого в нейронной сети.

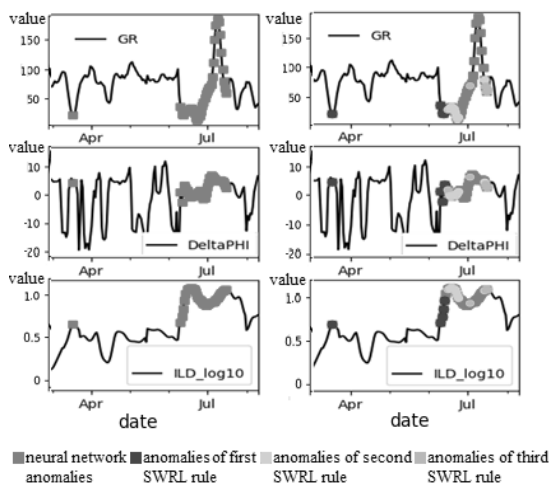


Рис. 7. Результат работы алгоритма детектирования аномалий

Класс и свойства, уже описанные в структуре онтологии, используются для загрузки индивидуумов в онтологию. Каждое значение временного ряда является отдельным индивидуумом, связанный с замером по этому показателю и замерами по другим показателям, зафиксированными в один момент. Онтология пересохраняется и правила запускаются после загрузки всех аномалий.

Все объекты онтологии могут быть извлечены из файла после обработки. В цикле проверяется флаг аномалии и выбираются все значения, принятые онтологией как истинно аномальные.

В левой части рисунка показаны все значения, определенные нейронной сетью как аномалии. Справа цвет указывает на данные, которые имеют метки в соответствии с разными правилами SWRL. Каждое из правил использует разные показатели из GR, Delta PHI и ILD. Разделение по нескольким правилам также позволяет классифицировать аномалии по категориям.

Мера средней абсолютной ошибки (MAE) использовалась для оценки качества модели. Система проверялась на 10 различных наборах данных одной предметной области. В среднем значение точности достигает 80%, что может свидетельствовать о том, что модель выявляет большинство отклонений и обрабатывает эти значения онтологией. Это значение не высокое. Нейронные сети этого типа достигают точности значительно более 90%. Однако использование только нейронных сетей не дает описанных выше преимуществ. Достижение наивысшей точности возможно за счет улучшения или замены архитектуры нейронной сети. Качество онтологии зависит от полноты используемых правил и не требует доработки структуры системы.

Произведено сравнение работы с методами KMeans и DBSCAN. На примерах более простых временных рядов математические алгоритмы показывают лучшие результаты. Но при появлении периодичности во временных рядах и усложнении паттернов временных рядов нейронные сети добиваются лучших результатов.

### Заключение

Таким образом, разработанный алгоритм поиска и обнаружения аномалий ВР, интегрирующий подход анализа ВР с помощью нейронных сетей и онтологический подход, показал результат порядка 80% эффективности на полученных тестовых данных работы нефтяных вышек. Данное значение не является максимальным, однако, архитектура алгоритма позволяет учесть экспертные мнения, формализованные в виде предметной онтологии.

Дальнейшими планами по развитию данного подхода поиска и обнаружения аномалий является усложнение структуры нейронной сети для получения наиболее оптимальных результатов.

Для онтологии улучшением является добавление классов метрик, которое позволит оперировать переменными, не привязанным к названиям конкретных параметров, а имеющие связи с классами метрик. В следствие этого расширение системы будет заключаться только в добавлении новых классов метрик и написании новых правил.

### Список литературы

- [Ahmad, 2016] Ahmad S., Purdy S., Real-time Anomaly Detection for Streaming Analytics, Numenta, Redwood City, CA, 2016.
- [Busseti, 2012] Busseti E., Osband I., Wong S. Deep Learning for Time Series Modeling, Stanford University, Stanford, California, 2012.
- [Chandola, 2007] Chandola V., Banerjee A., Kumar V., Anomaly Detection: a Survey, University of Minnesota, Minneapolis, MN, 2007.

- [**Das, 2009**] Das K. Detecting Patterns of Anomalies, Carnegie Mellon University, Pittsburgh, PA, 2009.
- [**Hasanov, 2018**] Hasanov V.I. Detection of anomalies in network traffic based on neural network modeling of dynamics of changes in IP packet volumes // MMC. – 2018. – No. 2.
- [**Jean-Baptiste, 2021**] Jean-Baptiste, L.: Ontologies with python. – Apress, Berkeley, CA 2021.
- [**Krizhevsky, 2010**] Krizhevsky A. Convolutional Deep Belief Networks on CIFAR-10, University of Toronto, Toronto, ON, 2010.
- [**Lindigrin, 2021**] Lindigrin, A.N. Analysis of specifics and problems of anomaly search processes in network data // Izvestiya TulaSU. Technical sciences. – 2021. – No. 5. – P. 304-309.
- [**Liu, 2016**] Liu X., Nielsen P. S. Regression-based Online Anomaly Detection for Smart Grid Data, Technical University of Denmark, Kgs. Lyngby, Denmark, 2016.
- [**Malhotra, 2015**] Malhotra, P., Vig, L., Shroff, G.M., & Agarwal, P. Long Short Term Memory Networks for Anomaly Detection in Time Series. ESANN. 2015.
- [**Moshkin, 2022**] Moshkin V., Yarushkina N., Moshkina I. The Intelligent System for Interactive Analysis and Forecasting of Graph Data / In: Kahraman C., Tolga A.C., Cevik Onar S., Cebi S., Oztaysi B., Sari I.U. (eds) // Intelligent and Fuzzy Systems. INFUS 2022. Lecture Notes in Networks and Systems. – Vol. 504. – Springer, Cham, 2022. – P. 870-878.
- [**Nikolov, 2010**] Nikolov T. Recurrent neural network based language model, Brno University of Technology, Brno, Czech Republic, 2010.
- [**Sak, 2014**] Sak H., Senior A., Beaufays F. Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling, Google Inc., Mountain View, CA, 2014.
- [**Wen, 2019**] Wen T., Keyes R. Time series anomaly detection using convolutional neural networks and transfer learning //arXiv preprint arXiv:1905.13628. – 2019.
- [**Yarushkina, 2016**] Yarushkina N., Moshkin V., Klein V., Andreev I, Beksaeva E.: Hybridization of Fuzzy Inference and Self-learning Fuzzy Ontology-Based Semantic Data Analysis. In: Proceedings of the First International Scientific Conference “Intelligent Information Technologies for Industry” (IITI'16). – 2016. – P. 277-285. – [https://doi.org/10.1007/978-3-319-33609-1\\_25](https://doi.org/10.1007/978-3-319-33609-1_25).
- [**Yarushkina, 2019**] Yarushkina N., Andreev I., Moshkin V., Moshkina I. Integration of Fuzzy OWL Ontologies and Fuzzy Time Series in the Determination of Faulty Technical Units / In: Misra S. et al. (eds) // Computational Science and Its Applications – ICCSA 2019. Lecture Notes in Computer Science. – 2019. – Vol. 11619. – Springer, Cham. – P. 545-555. – [https://doi.org/10.1007/978-3-030-24289-3\\_40](https://doi.org/10.1007/978-3-030-24289-3_40).
- [**Zhong, 2020**] Zhong C., Gursoy M. C., Velipasalar S. Anomaly Detection and Sampling Cost Control via Hierarchical GANs //GLOBECOM 2020-2020 IEEE Global Communications Conference. – IEEE, 2020. – P. 1-6.
- [**Zuev, 2021**] Zuev, V.N. Detection of network traffic anomalies by deep learning method // Software products and systems. – 2021. – No. 1. – P. 91-97.

УДК 510.642

## О СИСТЕМАХ СЛОЁВ В СТРУКТУРЕ СУПЕРИНТУИЦИОНИСТСКИХ ПРЕДИКАТНЫХ (ИЛИ ПРОПОЗИЦИОНАЛЬНЫХ) ЛОГИК. I

Д.П. Скворцов (*skvortsovd@yandex.ru*)  
Федеральный исследовательский центр  
«Информатика и управление» РАН, Москва

Рассматривается структура (решётка) суперинтуicionистских предикатных логик. Обсуждается подход к классификации логик этой структуры, то есть разбиению структуры на систему слоёв, задаваемых надлежащим отношением эквивалентности. Описаны (или упомянуты) некоторые примеры.

**Ключевые слова:** суперинтуicionистские предикатные логики, суперинтуicionистские пропозициональные логики, отношения эквивалентности и разбиения.

### Введение

Представляемый текст, если говоря формально, относится к области рассмотрения неклассических, точнее, суперинтуicionистских логик (в первую очередь, предикатных, и в меньшей степени, пропозициональных). Тем не менее, по ощущению автора, существо предмета выходит за рамки такой области, и скорее, связано с более общим подходом к обсуждению логических систем, к классификации таких систем на базе соотношений между их дедуктивными и семантическими характеристиками (то есть к представлению и описанию посредством логико-математических исчислений и определённого типа моделей, и к сопоставлению этих двух направлений взгляда на рассматриваемые сущности). Однако, чтобы избежать различных подстерегающих соблазнов и опасностей, чтобы не растекаться мыслию по, автор всё же предпочитает оставаться в близкой ему сфере обсуждения суперинтуicionистских логических систем<sup>1</sup>.

---

<sup>1</sup> Автор (возможно, не без воздействия персональных вкусовых предпочтений) не склонен возражать против мнения специалистов, кто полагают, что неклассические логики (в т.ч. суперинтуicionистские) могут быть интересным (а порой, и полезным) средством для интеллектуального анализа данных и рассуждений.



## 1. О структуре суперинтуиционистских логик

Чтобы не заморачиваться возникающими техническими деталями, навороченными привходящими обстоятельствами и осложнениями, ограничимся предикатным языком первого порядка, без равенства и функциональных символов, причём счётным (то есть с не более чем счётным количеством символов); иначе сказать, занимаемся логиками в, как когда-то говорили А. Чёрч и другие классики математической логики: узком предикатном языке.

Под *суперинтуиционистскими логиками* (предикатными или пропозициональными) понимаются (сравн. [Gabbay, etc., 2009; Sect.2.6, Definition 2.6.3]) произвольные множества формул (соответствующего вида), содержащие все аксиомы интуиционистского исчисления и замкнутые относительно традиционных правил вывода: в предикатном случае это *modus ponens*, правило обобщения (или правила Бернайса, если читатель их предпочитает) и правило подстановки произвольных формул вместо предикатных символов. В пропозициональном случае, исходя из того, что пропозициональные формулы – это частный случай формул предикатных, но без индивидуальных переменных (а значит, и без кванторов, и без предикатных символов арностей  $>0$ , так что в них остаются только символы 0-местные, то есть пропозициональные), то кванторное правило обобщения опускается, и подстановка применяется, естественно, пропозициональная. Говоря о логиках, будем систематически опускать слово ‘суперинтуиционистская’ (ну разве что изредка, ради напоминания), а также слово ‘предикатная’ (в знак того, что именно они – основной предмет нашего интереса), слово ‘пропозициональная’ можем опустить лишь если оно непосредственно восстанавливается из контекста, не вызывая коллизии недопонимания, или когда неважно, о каких именно логиках речь: сказанное применимо и к тем, и к другим. Логика  $\mathbf{L}$  понимается как множество её теорем, то есть доказуемых в ней формул, и вместо принадлежности логике ( $A \in \mathbf{L}$ ) порой говорим о доказуемости в  $\mathbf{L}$  и пишем  $\mathbf{L} \vdash A$ .

Наименьшее расширение логики  $\mathbf{L}$ , содержащее формулу  $A$ , обозначается  $[\mathbf{L}+A]$ ; для расширения логики произвольным множеством формул (дополнительных аксиом)  $\Gamma$  можно применять обозначение  $[\mathbf{L}+\Gamma]$ . Формулы  $A$  и  $A'$  назовём  *$\mathbf{L}$ -равносильными* (или *равносильными в логике  $\mathbf{L}$* ), если  $\mathbf{L} \vdash (A \leftrightarrow A')$ ; при этом, по правилу *modus ponens* обеспечивается, что:  $A \in \mathbf{L}' \Leftrightarrow A' \in \mathbf{L}'$  для всякого расширения  $\mathbf{L}'$  логики  $\mathbf{L}$ . В частности, если формулы *QH*-равносильны, то в произвольной (суперинтуиционистской) логике  $\mathbf{L}$  они доказуемы одновременно.

Известно, что совокупность  $\mathcal{A}$  всех логик (того или иного рода) образует структуру (то есть решётку, в традиционном теоретико-решёточном смысле), частично упорядоченную по включению, с операциями теорети-

ко-множественного пересечения и логической суммы (которая есть дедуктивное замыкание, относительно уже названных правил, для объединения логик). Наименьшей логикой в этой структуре, является, естественно, сама интуиционистская логика (обозначаемая  $\mathbf{H}$  в пропозициональном случае и  $\mathbf{QH}$  в предикатном), а наибольшей – (абсолютно) противоречивая логика  $\mathbf{L}^\perp = [(\mathbf{QH} + \perp)]$  (где  $\perp$  – пропозициональная константа “ложь”), то есть множество всех формул. Известно, что структура логик  $\mathbf{A}$  полна, то есть операции пересечения и логической суммы применимы к произвольной совокупности логик, конечной или бесконечной; впрочем, любая сумма логик на деле сводится к сумме счётного количества логик (поскольку, как уже сказано, предполагается, что число символов логического языка, а значит, и количество формул, счётно) – соответственно, уходить в проблемы, связанные с большими кардиналами, не придётся.

Известно, что наибольшей непротиворечивой суперинтуиционистской пропозициональной логикой является классическая логика

$$\mathbf{C} = [ \mathbf{H} + p \vee \neg p ] = [ \mathbf{H} + \neg \neg p \rightarrow p ]$$

(где  $p$  – пропозициональный символ), а все непротиворечивые пропозициональные логики, т.е. логики, заключённые между интуиционистской  $\mathbf{H}$  и классической  $\mathbf{C}$ , называют промежуточными.

С другой стороны, классическая предикатная логика

$$\mathbf{QC} = [ \mathbf{QH} + p \vee \neg p ] = [ \mathbf{QH} + \neg \neg p \rightarrow p ]$$

(ввиду теоремы Гёделя о полноте плюс теорема Лёвенгейма – Скулема, для произвольной бесконечной области  $D$  логика  $\mathbf{QC}$  есть множество формул, классически истинных в  $D$  – и значит, она же есть множество формул, классически истинных во всех (непустых) областях) – так вот, эта логика, в отличие от пропозиционального случая, имеет бесконечно много непротиворечивых расширений. А именно, для начала, это бесконечная убывающая  $\omega$ -цепь  $\mathbf{QC}_1 \supset \mathbf{QC}_2 \supset \dots \supset \mathbf{QC}_m \supset \mathbf{QC}_{m+1} \supset \dots$

логик  $\mathbf{QC}_m$  классической истинности в  $m$ -элементной области (для натуральных  $m > 0$ ), завершаемая их пересечением  $\mathbf{QC}_\omega = \bigcap \mathbf{QC}_m$  (логика классической истинности во всех конечных областях, или логика Трахтенброта, установившего, что она, в отличие от классической  $\mathbf{QC}$  и всех конечных  $\mathbf{QC}_m$ , не рекурсивно аксиоматизируема – а на самом деле и более того,  $\Pi_1$ -полна), плюс бесконечное ‘облако’ семантически не задаваемых логик, заключённых между логиками  $\mathbf{QC}$  и  $\mathbf{QC}_\omega$  (это ‘облако’ не может быть конечным, так как тогда и все его логики, и сама логика  $\mathbf{QC}_\omega$  оказались бы конечно аксиоматизируемы – вопреки результату Трахтен-

брота). Опять же, предикатные логики, заключённые между интуиционистской **QH** и классической **QC**, по традиции называют *промежуточными* – но теперь это уже отнюдь не все непротиворечивые суперинтуиционистские предикатные логики. А наибольшей непротиворечивой суперинтуиционистской предикатной логикой оказывается, что достаточно естественно, логика **QC**<sub>1</sub> (классическая логика одноэлементной области).

## 2. Системы слоёв

Говорят, что подсемейство  $S$  структуры  $A$  (всех) суперинтуиционистских логик (предикатных или пропозициональных) *выпукло* в  $A$ , если оно удовлетворяет такому условию:

$$\forall L_1, L_2 \in S \forall L_0 \in A [ (L_1 \subseteq L_0 \subseteq L_2) \Rightarrow (L_0 \in S) ].$$

Пустое семейство логик тривиальным образом выпукло, но нас будут интересовать только непустые выпуклые семейства. Начнём с очевидных примеров.

*Сегмент* (ограниченное выпуклое множество) есть полная подрешётка в  $A$  такого вида:  $[L_1, L_2] = \{ L \mid L_1 \subseteq L \subseteq L_2 \}$  (для таких логик  $L_1$  и  $L_2$ , что  $L_1 \subseteq L_2$ ). Естественно, *интервал* (или открытый интервал) в  $A$  имеет такой вид:  $(L_1, L_2) = \{ L \mid L_1 \subset L \subset L_2 \}$  (отметим, что условие  $L_1 \subset L_2$  необходимо, но не достаточно для непустоты интервала  $(L_1, L_2)$ ). Разумеется, эти простые примеры далеко не исчерпывают понятие выпуклости: скажем, любая антицепь (в структуре  $A$ ) выпукла. Этот радикальный пример иллюстрирует, что выпуклые подмножества, вообще говоря, не обязаны быть подрешётками в  $A$ .

Ясно, что выпуклое подмножество  $\Lambda$  есть сегмент в том и только в том случае, когда оно содержит наибольший и наименьший элементы.

С другой стороны, подсемейство  $S$  структуры логик  $A$  выпукло тогда и только тогда, когда оно вместе с любыми своими двумя логиками  $L_1 \subset L_2$  целиком включает в себя заключённый между ними сегмент  $[L_1, L_2]$ .

Теперь введём, сформулируем основное понятие, предлагаемое и обсуждаемое в этой работе.

Под *системой слоёв* ( $S_\theta$ ;  $\theta \in \Theta$ ) в структуре логик  $A$  понимаем разбиение  $A$  на выпуклые подмножества  $S_\theta$  (называемые, естественно, *слоями* данного разбиения). Это означает, что все слои непусты, дизъюнкты и в объединении составляют всю структуру  $A$ . Известно, что *каждое такое разбиение задаётся соответствующим отношением эквивалентности на структуре  $A$* .

Разумеется, подобная система слоёв заслуживает внимания, обсуждения и рассмотрения, только если само отношение эквивалентности на  $\mathcal{A}$  или порожаемые им классы эквивалентности (то есть слои) чем-то интересны или осмысленны. Можно выразить надежду, что в наиболее привлекательных системах большинство слоёв (или хотелось бы, чтобы все?) окажутся сегментами – однако гарантировать этого изначально (т.е. до обретения достаточного практического опыта в работе с различными реально возникающими системами слоёв) никто не может<sup>2</sup>. Во всяком случае, более того, мы не готовы априорно запрещать или исключать систему слоёв только на том основании, что какие-то из них не окажутся подрешётками структуры  $\mathcal{A}$ . Хотя всё же не ожидаем, что понадобится рассматривать системы, где слои окажутся, положим, нетривиальными (т.е. не одноэлементными) антицепями.

Начнём с пары примеров. Автор надеется, что рассмотрение этих и иных им подобных (похожих или не очень похожих?) иллюстраций способно пролить свет на возможности и перспективы предлагаемого подхода к классификации неклассических логик (или других логических систем, а может быть, и не только их?) посредством исследования возникающих систем слоёв. Способно помочь увидеть и оценить его применимость и пользу для более глубокого и детального постижения и понимания строения соответствующей структуры (в нашем случае, структуры суперинтуиционистских логик).

### 3. Пропозициональные фрагменты предикатных логик

*Пропозициональным фрагментом*  $\pi[\mathbf{L}]$  предикатной (суперинтуиционистской) логики  $\mathbf{L}$  называют совокупность всех принадлежащих ей (то есть доказуемых в ней) пропозициональных формул (сравн. [Ono, 1972-1973; Sect.6] и [Gabbay etc., 2009; Sect.2.15]). Очевидно, этот фрагмент является пропозициональной (суперинтуиционистской) логикой (поскольку пропозициональная подстановка по существу есть частный случай предикатной подстановки).

Скажем, что две предикатные логики *пропозиционально эквивалентны* (или  *$\pi$ -эквивалентны*), если они имеют один и тот же пропозициональный фрагмент:  $(\mathbf{L}_1 \approx_{\pi} \mathbf{L}_2) \Leftrightarrow (\pi[\mathbf{L}_1] = \pi[\mathbf{L}_2])$ . Это есть отношение эквивалентности на структуре  $\mathcal{A}$  предикатных логик, порожающее классы эквивалентности  $(S_{\lambda} : \lambda \in \mathcal{A}_{\pi})$ , где  $\mathcal{A}_{\pi}$  здесь означает структуру всех пропозициональных суперинтуиционистских логик, а  $S_{\lambda} = \{\mathbf{L} \in \mathcal{A} \mid \pi[\mathbf{L}] = \lambda\}$  есть совокупность предикатных логик с заданным (фиксированным) пропозициональным фрагментом  $\lambda$  (для  $\lambda \in \mathcal{A}_{\pi}$ ); элементы  $S_{\lambda}$  называют *предикатными расширениями* пропозициональной логики  $\lambda$ .

---

<sup>2</sup> как известно, гарантии раздаёт только небезызвестный полис – да и тот неясно, навряд ли обеспечивает их на деле?

Ясно, что все эти классы  $S_\lambda$  выпуклы, то есть они образуют систему слоёв пропозициональной эквивалентности, или  $\pi$ -слоёв. Удачным образом оказывается, что все  $\pi$ -слои являются сегментами в структуре  $\Lambda$ , т.е.  $S_\lambda = \{ L \in \Lambda \mid L_\lambda^- \subseteq L \subseteq L_\lambda^+ \}$  для  $\lambda \in \Lambda_\pi$ . Здесь логики

$$L_\lambda^- = [QH + \lambda] \text{ и } L_\lambda^+ = [QH + \text{DOM}_1 + \lambda] = [QH + \text{DOM}_1' + \lambda]$$

называются, соответственно, *наименьшим* и *наибольшим предикатным расширением* пропозициональной логики  $\lambda$ . Вторая из них аксиоматизируется с использованием формул

$$\text{DOM}_1 = \exists x P(x) \rightarrow \forall x P(x) \text{ или } \text{DOM}_1' = \forall x \forall y (P(x) \rightarrow P(y)),$$

естественным образом выражающих, положим, в предикатной классической семантике (а впрочем, и в семантике Крипке для интуиционистской предикатной логики) свойство одноэлементности предметной области.

В частности,  $\pi$ -слой  $S_H$  наименьшей (т.е. интуиционистской) пропозициональной логики  $H$  есть сегмент, заключенный между наименьшей (т.е. опять же интуиционистской) предикатной логикой  $L_H^- = QH$  и логикой  $L_H^+ = [QH + \text{DOM}_1] = [QH + \text{DOM}_1']$ .

Аналогично,  $\pi$ -слой  $S_C$  классической пропозициональной логики  $C$  есть сегмент, заключенный между классической предикатной логикой  $L_C^- = QC$  и логикой  $L_C^+ = [QC + \text{DOM}_1] = [QC + \text{DOM}_1'] = QC_1$  (классическая логика одноэлементной области), т.е. этот слой есть в точности сегмент всех непротиворечивых расширений классической предикатной логики, описанный в заключительном абзаце раздела 0.

Наконец,  $\pi$ -слой противоречивой пропозициональной логики состоит из единственной логики: тоже противоречивой, но только теперь уже предикатной. Это – единственный ‘вырожденный’, то есть одноэлементный  $\pi$ -слой: в котором ‘нижняя’ и ‘верхняя’ его логики совпадают.

При этом сегмент всех расширений логики  $L_H^+ = [QH + \text{DOM}_1] = [QH + \text{DOM}_1']$ , т.е.  $\{ L \in \Lambda \mid L_H^+ \subseteq L \}$ , оказывается равен  $\{ L_\lambda^+ \mid \lambda \in \Lambda_\pi \}$ , т.е. он состоит в точности из наибольших расширений  $L_\lambda^+$  всех суперинтуиционистских пропозициональных логик. Таким образом, имеется естественный изоморфизм между этим сегментом предикатных логик и структурой  $\Lambda_\pi$  всех пропозициональных логик. С алгоритмической точки зрения этот изоморфизм есть 1-эквивалентность, и значит, он сохраняет многочисленные ‘разумные’ алгоритмические характеристики, такие как разрешимость, рекурсивная аксиоматизируемость и другие. Также он сохраняет естественные семантические свойства логик, как например, полноту (для алгебраической семантики или для семантики Крипке), а также финитную аппроксимируемость (которая означает, что логика задаётся

своими конечными шкалами: равносильно, в алгебраической или в Крипке семантике). Отсюда, в частности, автоматически вытекает, что все предикатные логики из обсуждаемого сегмента (расширений  $\mathbf{L}_H^+$ ) оказываются алгебраически полны (поскольку известно, что все суперинтуиционистские пропозициональные логики таковы), а скажем, сама логика  $\mathbf{L}_H^+ = [ \mathbf{QH} + \text{DOM}_1 ] = [ \mathbf{QH} + \text{DOM}_1' ]$  Крипке-полна, и более того, финитно аппроксимируема (ведь такова соответствующая ей интуиционистская пропозициональная логика  $\mathbf{H}$ ), и тому подобное.

Некоторые элементы этой общей картины установлены в уже процитированных работах, а целостное её изложение (включая, возможно, иные, более технические аспекты и моменты, которых здесь касаться не станем) может быть представлено в последующих работах.

#### 4. Негативные фрагменты предикатных логик

*Негативным фрагментом*  $\text{Neg}[\mathbf{L}]$  предикатной (суперинтуиционистской) логики  $\mathbf{L}$  назовём совокупность всех принадлежащих ей (т.е. доказуемых в ней) формул вида  $\neg A$ . Введём также  $\neg\neg$ -фрагмент

$\mathbf{L}^{\neg\neg} = \{ \neg\neg A \mid \neg\neg A \in \mathbf{L} \}$  (множество доказуемых в  $\mathbf{L}$  формул вида  $\neg\neg A$ ). Ясно, что  $\mathbf{L}^{\neg\neg}$  есть собственное подмножество  $\text{Neg}[\mathbf{L}]$  (т.е.  $\mathbf{L}^{\neg\neg} \subset \text{Neg}[\mathbf{L}]$  для всякой логики  $\mathbf{L}$ ).

При этом *оба фрагмента взаимно задают друг друга*: а именно,

$\mathbf{L}^{\neg\neg} = \{ \neg\neg A \mid \neg\neg A \in \text{Neg}[\mathbf{L}] \}$  и обратно  $\text{Neg}[\mathbf{L}] = \{ \neg A \mid \neg\neg\neg A \in \mathbf{L}^{\neg\neg} \}$

(так как  $\mathbf{QH} \vdash (\neg\neg\neg A \leftrightarrow \neg A)$ , т.е. формулы  $\neg\neg\neg A$  и  $\neg A$   $\mathbf{QH}$ -равносильны, и значит, принадлежат произвольной логике одновременно).

Скажем, что две предикатные логики *негативно эквивалентны* (или короче, *Neg-эквивалентны*), если они имеют один и тот же негативный фрагмент (или, что равносильно, тот же  $\neg\neg$ -фрагмент):

$$(\mathbf{L}_1 \approx_{\text{Neg}} \mathbf{L}_2) \Leftrightarrow (\text{Neg}[\mathbf{L}_1] = \text{Neg}[\mathbf{L}_2]) \Leftrightarrow ((\mathbf{L}_1)^{\neg\neg} = (\mathbf{L}_2)^{\neg\neg}).$$

Это есть отношение эквивалентности на структуре  $\mathcal{A}$  предикатных логик, классы которого очевидным образом выпуклы, то есть оно задаёт систему *Neg-слоёв* в структуре предикатных логик  $\mathcal{A}$ . Эти слои можно обозначать как  $S_L$  для произвольной логики из данного слоя, или как  $S_{\text{Neg}[\mathbf{L}]}$  (благо  $\text{Neg}[\mathbf{L}]$  не зависит от выбора любой из числа Neg-эквивалентных логик). Опять, счастливым образом, *все Neg-слои оказываются сегментами в структуре  $\mathcal{A}$* , т.е.  $S_{\text{Neg}[\mathbf{L}]} = \{ \mathbf{L} \in \mathcal{A} \mid \mathbf{L}_{\text{Neg}[\mathbf{L}]}^- \subseteq \mathbf{L} \subseteq \mathbf{L}_{\text{Neg}[\mathbf{L}]}^+ \}$ .

Здесь, естественно,  $\mathbf{L}_{\text{Neg}[\mathbf{L}]}^- = [ \mathbf{QH} + \text{Neg}[\mathbf{L}] ] = [ \mathbf{QH} + \mathbf{L}^{\neg\neg} ] = [ \mathbf{QH} + \mathbf{L}^{\neg\neg'} ]$ , где  $\mathbf{L}^{\neg\neg'} = \{ \neg\neg A \mid A \in \mathbf{L} \}$  есть модифицированный (или урезанный)  $\neg\neg$ -фрагмент логики  $\mathbf{L}$  (последнее равенство логик обеспечивается тем, что если  $\neg\neg A \in \mathbf{L}^{\neg\neg}$ , то есть  $\neg\neg A \in \mathbf{L}$ , то  $\neg\neg\neg\neg A \in \mathbf{L}^{\neg\neg'}$ , а

формулы  $\neg\neg\neg\neg A$  и  $\neg\neg A$ , опять же, **QH**-равносильны). Очевидно, что  $L^{\neg\neg} \subseteq L^{\neg}$  для всякой логики **L** (так как **QH**  $\vdash (A \rightarrow \neg\neg A)$  для всякой формулы  $A$ ); более того, для всех логик, кроме классической **QC** и её расширений, это включение оказывается строгим, собственным (а именно, если **L** не содержит **QC**, то  $(\neg\neg p \rightarrow p) \notin L$ , а значит, интуиционистски доказуемая формула  $\neg\neg(\neg\neg p \rightarrow p)$  принадлежит разности  $(L^{\neg\neg} \setminus L^{\neg})$ ).

*Предупреждение:* Урезанный  $\neg\neg$ -фрагмент  $L^{\neg\neg}$ , на первый взгляд, привлекает тем, что он связывает  $\neg\neg A$  с принадлежностью к **L** самой формулы  $A$  – но возможно, как раз поэтому он, как вскоре увидим, не годится, не подходит для описания Neg-эквивалентности логик; фрагменты  $L^{\neg\neg}$  и  $L^{\neg}$  неразличимы только относительно логической доказуемости.

С другой стороны, удовлетворительное представление для ‘верхней’ логики  $L_{\text{Neg}[L]}^+$  в общем случае автору неизвестно: её можно задать как логическую сумму всех логик из данного слоя, а теорема дедукции для предикатных логик позволяет установить, что эта сумма принадлежит тому же слою (то есть Neg-эквивалентна всем суммируемым логикам), и значит, является в нём наибольшей (эти соображения предварительным образом рассматривались в пока не опубликованной, и даже ещё не подготовленной к печати работе [Skvortsov, 2018]).

Стоит отметить, что эта конструкция не переносится разумным образом на пропозициональный случай – точнее, там аналогичная система Neg-слоёв оказывается вырожденной и интереса не представляет. А именно, по известной *теореме Гливенко*, все классически доказуемые пропозициональные формулы доказуемы уже и интуиционистски, т.е. негативные фрагменты у логик **H** и **C**, а значит, и вообще у всех промежуточных пропозициональных логик совпадают. Поэтому здесь Neg-эквивалентность разбивает структуру пропозициональных логик на два класса, один из которых состоит из противоречивой логики, а другой собирает, вбирает в себя все остальные. Разумеется, подобная дегенеративная система слоёв ни малейшего ни рассмотрения, ни обсуждения не заслуживает.

В предикатном случае ситуация выглядит куда как более нетривиальной и остаётся малоисследованной – точнее, почти не исследованной. Первым *класс Neg-эквивалентности* (не вводя такого общего понятия) *рассмотрел Габбай в 1972 году* (см. [Gabbay, 1972]). А именно, он описал сегмент логик, имеющих классический негативный фрагмент (т.е. логик, в которых доказуемы те же формулы вида  $\neg A$ , что и в классической предикатной логике **QC**). Точнее, он установил (с применением той же предикатной теоремы дедукции), что классический негативный фрагмент имеет так называемая *логика Куроды*:

$$\mathbf{QK} = [\mathbf{QH} + \mathbf{K}] = [\mathbf{QH} + \mathbf{K}'] = [\mathbf{QH} + \mathbf{K}''],$$

где  $K = \forall x \neg\neg P(x) \rightarrow \neg\neg\forall x P(x)$ ,  $K' = \neg\neg\forall x (P(x) \vee \neg P(x))$ ,

$K'' = \neg\neg\forall x (\neg\neg P(x) \rightarrow P(x))$  (собственно говоря, Курода и нашёл, придумал свою формулу (или свою логику?) именно в качестве контрпримера к предикатному аналогу теоремы Гливенко: классически доказуемая, но интуиционистски недоказуемая предикатная формула вида  $\neg A$ ). Из этого немедленно вытекает, что классом Neg-эквивалентности (т.е. Neg-слоем) классической логики **QC** является сегмент  $S_{\text{Neg}[\text{QC}]} = [\mathbf{QK}, \mathbf{QC}]$  логик, заключённых между логикой Куроды **QK** и классической **QC**.

К слову, отметим, что формула  $\forall x (P(x) \vee \neg P(x))$  не доказуема в логике **QK** (в отличие от классической **QC**), а значит, формула  $K'$  принадлежит разности  $(\mathbf{QC}^{\neg\neg} \setminus \mathbf{QK}^{\neg\neg})$  урезанных  $\neg\neg$ -фрагментов Neg-эквивалентных логик **QC** и **QK** (таким образом, Neg-эквивалентность логик не означает совпадения их  $\neg\neg$ -фрагментов).

Недавно автору удалось описать аналогичные сегменты (Neg-слои) для классических логик  $\mathbf{QC}_m$  конечных,  $m$ -элементных областей (для  $0 < m < \omega$ ): а именно, каждый такой слой есть сегмент  $[\mathbf{QK}_m, \mathbf{QC}_m]$ , ограниченный *снизу логикой*  $\mathbf{QK}_m = [\mathbf{QK} + \neg\text{Dom}_m]$ , а *сверху* – естественно, *самой логикой*  $\mathbf{QC}_m$ . Здесь каждая  $\neg\text{Dom}_m$  есть сравнительно длинная, но идейно достаточно прозрачная формула; она есть в некотором естественном, точном смысле наименьшая, слабейшая формула, достаточная для аксиоматизации логики  $\mathbf{QC}_m$ , при добавлении к классической логике **QC** (см. [Skvortsov, 2023]).

Про негативный фрагмент логики Трахтенброта  $\mathbf{QC}_\omega$  удаётся сказать только, что он есть сегмент между ‘нижней’ логикой

$$[\mathbf{QH} + \text{Neg}[\mathbf{QC}_\omega]] = [\mathbf{QH} + \mathbf{QC}_\omega^{\neg\neg}] \text{ (здесь, ясное дело,}$$

$$\mathbf{QC}_\omega^{\neg\neg} = \mathbf{QC}_\omega^{\neg\neg'} = \{ \neg\neg A \mid A \in \mathbf{QC}_\omega \} )$$

и ‘верхней’ логикой: разумеется, *самой*  $\mathbf{QC}_\omega$ . Впрочем, *бóльшего* от логик **L** из этого Neg-слоя ждать и не стоит: *любая из них*  $\Pi_1$ -трудна, так как к ней сводима  $\Pi_1$ -полная логика  $\mathbf{QC}_\omega$  (а именно,  $(A \in \mathbf{QC}_\omega) \Leftrightarrow (\neg\neg A \in \mathbf{L})$ ). Для читателя, любопытствующего положением, местом числовых множеств (в данном случае, логик) в арифметической иерархии, можно отметить, что ‘*нижняя*’ логика заведомо *принадлежит* классу  $\Sigma_2$  (как аксиоматизированная  $\Pi_1$ -множеством формул  $\mathbf{QC}_\omega^{\neg\neg}$ ).

С другой стороны, Neg-слои остальных суперинтуиционистских логик, похоже, остаются неизученными и неописанными. Например, автору неизвестны собственные расширения интуиционистской логики **QH** с интуиционистским негативным фрагментом. Не исключено, Neg-слой интуиционистской логики может состоять из одной этой логики, но и это остаётся непроверенным.



Быть может, заслуживали бы рассмотрения Neg-слои таких известных логик как:

(I)  $QHJ = [QH + J]$  (наименьшее предикатное расширение пропозициональной логики слабого закона исключенного третьего  $HJ = [H + J]$ ), где  $J = \neg\neg p \vee \neg p$ ,

(II)  $QLC = [QH + Z]$  (аналогичное расширение для логики Даммета или логики цепей  $LC = [H + Z]$ ), где  $Z = (p \rightarrow q) \vee (q \rightarrow p)$ , и для различных иных. Например, предстоит проверить, достаточны ли для аксиоматизации ‘нижних’ логик соответствующих слоёв такие формулы:

для (I): формулы

$\neg\neg\forall x_1, \dots, x_n (\neg\neg P(x_1, \dots, x_n) \vee \neg P(x_1, \dots, x_n))$  со всеми натуральными  $n$ ;

для (II): формулы

$\neg\neg\forall x_1, \dots, x_n [(P(x_1, \dots, x_n) \rightarrow Q(x_1, \dots, x_n)) \vee (Q(x_1, \dots, x_n) \rightarrow P(x_1, \dots, x_n))]$ .

Если эта гипотеза верна, то возникает естественный **вопрос**: конечно аксиоматизируемы ли получаемые ‘нижние’ логики?

Автор подозревает, что ответ на него вполне может оказаться отрицательным.

Кроме того, в отличие от классических логик  $QC$  и  $QC_m$  (для  $m \leq \omega$ ), здесь не кажется слишком правдоподобным предположение, что ‘верхними’ логиками обсуждаемых Neg-слоёв окажутся сами логики  $QHJ$  или  $QLC$  (и т.п.), так что вопрос обозримого, пристойного описания этих ‘верхних’ логик остаётся открытым, и может оказаться нетривиальным.

Короче, в обсуждаемой области видятся некоторые небезлюбопытные вопросы, но кавалерийским наскоком иные из них не берутся, а ждут более обстоятельного приступа и обдумывания.

## Заключение. Перспективы

В продолжении, второй части представленной работы (которая может быть подготовлена и опубликована в будущем) автор надеется обсудить другие возможные классификации, системы слоёв, возникающие (или скорее, имеющие место быть, но просто замечаемые, обнаруживаемые, выявляемые) на структуре предикатных (или пропозициональных) суперинтуиционистских логик.

В числе их: так называемая *иерархия конечных слоёв* (называемых

*FD-слоями*: finite domains based slices, слои на базе конечных областей, см. [Skvortsov, 2023]). Эта система слоёв порождается сегментом суперклассических логик (расширений классической  $QC$ ), описанным в заключительном абзаце раздела 0, и параметризована ординалами  $m \leq (\omega + 1)$ ; она представляет собой аналог (не исключено, один из многих?) иерархии конечных слоёв, которую описал Хосои в 1960-х годах для структуры су-

перинтуитионистских пропозициональных логик (см., напр., [Gabbay etc., 2009; Sect.1.15 ]). Не исключено, в докладе автор мог бы попытаться слегка наглядно представить эту иерархию слоёв, в комбинации с (в некотором определённом смысле) ‘перпендикулярной’ ей системой  $\pi$ -слоёв, описанной в разделе 2.

Другая возможная система слоёв (вероятно, очень тяжёлая для исследования?) связана с семантикой Крипке: с понятиями полноты, неполноты, пополнения – и вдаваться в необходимые для этого технические детали сейчас не станем.

Также в будущем можно рассмотреть и иные системы слоёв, связанные с разными ограниченными фрагментами предикатного языка: подобно системе  $\pi$ -слоёв, где речь идёт о фрагментах пропозициональных, то есть без обоих кванторов. А можно аналогично использовать фрагменты без одного квантора: всеобщности или существования. Или тому подобное.

Результаты работы могут также оказаться полезны при развитии логики предикатов для эмпирических модальностей, и это может быть использовано для формализации эмпирических закономерностей в интеллектуальных системах [Финн, 2022].

**Благодарности.** Автор выражает благодарность В.Б. Шехтману, а также анонимным рецензентам за полезные соображения, способствовавшие и содействовавшие усовершенствованию изложения. Также автор крайне признателен М.А. Михеенковой за неоценимую помощь в подготовке ‘публикабельной’ версии здешнего текста.

## Список литературы

- [Финн, 2022] Финн В.К. ДСМ-рассуждения и обнаружение знаний: ампликативные выводы, распознавание причинности и три вида полноты // Научно-техническая информация. Сер. 2. – 2022. – № 4.
- [Gabbay, 1972] Gabbay D. Application of trees to intermediate logics // Journal of Symbolic Logic. – 1972. – No. 37.
- [Gabbay etc., 2009] Gabbay D., Shehtman V., Skvortsov D. Quantification in nonclassical logic. Vol. 1. Sections 1.15, 2.6, 2.15 // Studies in Logic and the Foundations of Mathematics. – Vol. 153. – Elsevier, 2009.
- [Ono, 1972-1973] Ono H. A study of intermediate predicate logics. Publications of RIMS, Kyoto Univ. – 1972-1973. – Vol. 8, No. 3.
- [Skvortsov, 2018] Skvortsov D. On the negative fragments of superintuitionistic predicate logics. – 2018. In preparation.
- [Skvortsov, 2023] Skvortsov D. On finite domains based slices in the structure of superintuitionistic predicate logics, Preview // Logical Investigations. To appear in 2023.

## ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ РИСКОВ НА ПРИМЕРЕ ЭПИДЕМИИ COVID-19\*

А.В. Соколов (*alexander.v.sokolov@gmail.com*)

Институт проблем передачи информации  
им. А.А. Харкевича РАН, Москва

Г.В. Ройзензон (*rgv@isa.ru*)

Институт системного анализа ФИЦ ИУ РАН, Москва

Н.П. Комендантова (*komendan@iiasa.ac.at*)

Л. Экенберг (*lovek@dsv.su.se*)

Международный институт прикладного  
системного анализа, Лаксенбург, Австрия

В работе предложена схема построения единой многомерной классификации методов интеллектуального анализа рисков. В рамках модельного подхода анализа рисков используется технология сбалансированной идентификации. Указанная технология послужила основой для создания системы мониторинга и прогноза состояния опасных явлений и объектов. Рассмотрен практический пример использования предложенного подхода для прогнозирования развития эпидемии Covid-19 в г. Москва (с марта 2020 г. по сентябрь 2022 г.). Показано, как расхождение прогноза и реальности приводит (после критического анализа) к модификации модели или пересмотру принятого сценария внешнего воздействия. Критически проанализированы перспективы развития новых методов интеллектуального анализа рисков.

**Ключевые слова:** моделирование, сбалансированная идентификация, управление рисками, искусственный интеллект.

### Введение

Пандемия коронавируса (далее CV19) продемонстрировала неготовность как многих стран (включая развитые), так и различных крупных международных организаций (ООН, ВОЗ и др.) противостоять новым уг-

---

\* Исследование выполнено при финансовой поддержке РФФИ и ИИАСА в рамках научного проекта № 20-57-82004.

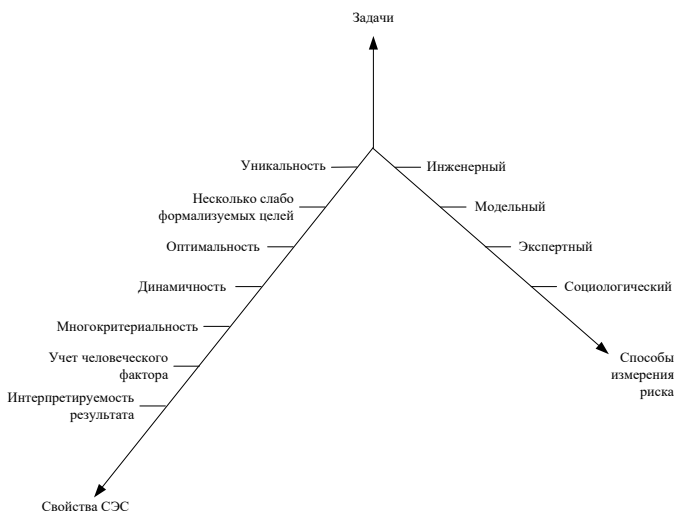
розам, а также формулировать и оценивать новые типы рисков [Черешкин и др., 2020]. В работе предложен новый и усовершенствован имеющийся математический инструментарий, ориентированный на решение различных задач анализа рисков и прогнозирования. Особое внимание требует разработка новых методов оценки эффективности при введении различных социальных ограничений для успешного противостояния пандемии CV19 (вакцинация, самоизоляция, мониторинг пассажиропотоков, перевод на удаленную форму работы и др.). Современные реалии требуют, наряду с классическими методами, использования для решения поставленных задач технологий искусственного интеллекта (далее ИИ).

## 1. Методология

Методы анализа рисков (далее MAP) можно условно разделить на четыре большие группы [Черешкин и др., 2020]. Первая группа – вероятностные (или инженерные) методы. В рамках данного подхода основные усилия направлены на сбор статистических данных о поломках, авариях и т.п. Инженерные методы ориентированы на количественный расчет вероятности поломок, отказов и других нежелательных событий [Королев и др., 2011]. Вторая группа MAP – построение моделей на основе данных (data driven modeling) [Геловани и др., 2018]. Данный подход предполагает моделирование процессов, определяющих динамику системы, которая может попасть в различные нежелательные состояния (аварии, эпидемии и т.п.) [Sokolov et al., 2020], [Соколов и др., 2021]. При этом важным этапом является выбор модели, соответствующей количеству и качеству исходной информации (цифровым данным и знаниям о функционировании моделируемого объекта). Для этого предлагается использовать метод сбалансированной идентификации [Sokolov et al., 2020], позволяющий достигнуть компромисса между сложностью математической модели и погрешностью описания используемого массива данных. Для использования метода необходимо: 1) определить (параметрическое) семейство моделей, пригодных для удовлетворительного воспроизведения измерений; 2) формализовать понятие сложности модели (для выбранного семейства); 3) задать меру близости траектории модели к исходным данным; 4) определить процедуру оценки погрешности моделирования (например, использовать процедуру перекрестной проверки); 5) провести поиск оптимального компромисса между сложностью модели и близостью к измерениям на основе минимизации погрешности моделирования измерений (перекрестная проверка). Кроме того, авторами работы на протяжении длительного периода развивается оригинальный подход глобального моделирования, основанный на использовании сразу нескольких моделей (мультимодельный подход) [Геловани и др., 2018], [Бритков и др., 2018]

применение которого для решения поставленной задачи представляется весьма перспективным. Мульти모델ный подход использован при решении различных крупномасштабных практических задач (например, оценки экономических и демографических потенциалов стран) [Геловани и др., 2018]. Третья группа МАР – экспертные. При применении инженерного и модельного подходов достаточно часто возникают ситуации, когда наблюдается дефицит статистических данных (или есть сомнения в их достоверности) [Кобринский и др., 2022], [Талеб, 2018]. Кроме того, при построении моделей в ряде случаев затруднительно выявить различные зависимости (слабоструктурированные задачи). В такой ситуации фактически единственным источником сведений являются эксперты. В рамках экспертного подхода хорошо себя зарекомендовали методы вербального анализа решений [Ларичев, 2006], [Асанов и др., 2001], [Ройзензон, 2012] ориентированные на слабоструктурированные задачи многокритериального стратегического выбора. Экспертные методы могут быть непосредственно использованы как для оценки рисков [Асанов и др., 2001], так и для оценки эффективности от введения дополнительных ограничительных мер, что предполагает разработку специальных систем критериев, по которым можно будет судить о степени достижения поставленных задач (целей) [Соколов и др., 2021]. При этом критерии могут быть условно разделены на три большие группы. Первую группу образуют критерии, позволяющие для противодействия пандемии CV19 оценивать имеющийся ресурс (масштабность проведения вакцинации и тестирования населения, количество больничных коек, приборов ИВЛ, медицинского персонала, медикаментов и т.п.). Вторая группа критериев позволяет оценивать скорость расходования и прироста ресурса во времени (возможность использования медицинских специалистов смежных специальностей, ускоренный ввод в эксплуатацию объектов медицинской сферы, возможности закупки лекарств и оборудования за рубежом и др.), необходимого для борьбы с пандемией CV19. Наконец, третья группа критериев позволяет сделать вывод о степени достижения поставленных целей (например, предполагается, что в течение определенного срока (например, года) вирусом заразится не более 2 процентов населения, или процент падения ВВП не превысит заданный порог и т.п.). Таким образом, оценки по указанным составным критериям [Ройзензон, 2005], [Ройзензон, 2012], позволяют сделать вывод, насколько принятые меры и полученные результаты, являются эффективными. Кроме того, важнейшим направлением использования экспертного подхода для анализа риска является возможность исследования вопросов безопасности критических инфраструктур (например, приостановка работы системообразующих предприятий, транспортных систем и т.п.) в условиях пандемии CV19 [Черешкин и др.,

2020]. Четвертый подход MAP – социологический [Бек, 2000]. В рамках данного подхода предполагается измерить восприятие населением и его отдельными группами того или иного риска. В рамках перспективного направления исследований предполагается провести дополнительный анализ влияния человеческого фактора [Смирнов и др., 2022], [Furems et al., 2003], [Смолян и др., 2002], [Поспелов и др., 1990], [Канеман и др., 2021], [Лефевр, 2003] на безопасность критически важных объектов и сформулировать рекомендации по его учету при расчете и анализе рисков в условиях пандемии CV19.



*Рис. 1. Многомерный классификатор методов анализа риска*

Указанные способы измерения риска положены в основу разработки многомерного классификатора интеллектуальных методов анализа риска, который был предложен авторами настоящей работы [Черешкин и др., 2020] (см. рис. 1). Разработанный классификатор интеллектуальных методов анализа риска может рассматриваться как удобный аналитический инструмент, т.к. позволяет выявлять «пробелы» в области разработки указанных методов. Подобный подход дает возможность для выбранных конкретных задач, с учетом свойств различных социально-экономических систем, а также способов измерения риска обозначить перечень уже разработанных методов или методов, разработка которых и является перспективной проблемой. Особенностью предложенного подхода является разработка методологии, которая может быть положена в основу создания

общей теории риска, гармонично синтезирующей как два основных направления в области ИИ (методы, основанные на знаниях и данных), так и четыре направления измерения риска (инженерный, модельный, экспертный и социологический). Предложенный комбинированный подход дает возможность эффективно сопоставлять и верифицировать полученные результаты, например, выявлять общие биологические закономерности и социальные особенности управления пандемией CV19 в различных регионах (странах). Еще одной важной фундаментальной научной задачей, которая решается в представленной работе, является разработка новых многокритериальных методов оценки эффективности при введении различных социальных ограничений для успешного противостояния пандемии CV19.

## 2. Моделирование

Эффективное проведение мониторинга состояния и прогноза поведения опасных явлений или объектов предполагает широкое использование математических моделей [Романюха, 2012], [Brauer et al., 2019], [Sokolov et al., 2020], [Попков и др., 2021], [Rahimi et al., 2021]. Для решения задачи математического моделирования используются различные подходы, а именно: регрессионные модели [Cook et al., 1982], методы машинного обучения [Попков и др., 2021], [Tamang, et al., 2020], генетические алгоритмы [Salgotra et al., 2021], [Гладков и др., 2006], модели SIR [Kermack et al., 1927], [Brauer et al., 2019] и ряд других [Rahimi et al., 2021]. Основная проблема применения на практике вышеперечисленных подходов моделирования является трудность в интерпретации (в современной терминологии – объяснительный ИИ). Цели мониторинга и прогноза ставятся с учетом наличия математических моделей, сложность, подробность и надежность которых в свою очередь определяются количеством и качеством (точностью) экспериментальных данных (наблюдений) и знаний о функционировании объекта.

Иногда построение системы мониторинга и прогноза начинается с нуля и по мере накопления данных и знаний и построения все более подробных и точных моделей возможности мониторинга и прогноза возрастают. В доступных рядах данных проявляется все больше эффектов, что позволяет усложнять соответствующие математические модели. Такому усложнению способствует и появление новых знаний об исследуемых процессах, которые иногда могут частично заменить цифровые данные.

Динамика эпидемии определяется процессами взаимодействия между вирусом, человеческим организмом и обществом. Различные процессы характеризуются неопределенными временными интервалами, причем, чем больше характерное время процесса, тем больше времени требуется для его проявления, тем длиннее должны быть ряды наблюдений, необходимые для определения его характеристик.

Для пандемии CV19 характерные времена различных процессов могут быть оценены следующим образом:

15 дней – заразность и манифестация (выявляемость) как функции длительности заболевания;

30 дней – текущий индекс репродукции (контактное число) и индекс выявления (и последующей изоляции);

60 дней – зависимость индекса выявления от количества проведенных тестов;

90 дней – влияние естественного (после болезни) коллективного иммунитета;

200 дней – влияние ослабления естественного иммунитета со временем;

100 дней – влияние вакцинации;

180 дней (предварительная оценка) – влияние ослабления искусственного (после вакцинации) иммунитета;

365 дней – сезонность заразности;

180 дней – появление новых штаммов.

Горизонт прогноза определяется погрешностью модели, которая характеризуется точностью описания отдельных процессов, определяемой, в свою очередь, длиной рядов наблюдений (и их надежностью). Таким образом, длительность мониторинга определяет, какие процессы могут быть оценены (с характерными временами меньше этой длительности) и, следовательно, определяют горизонт и точность прогноза. С этой точки зрения становится понятной неудача попытки прогноза динамики пандемии, принятой Н. Фергюсоном. Его группа в начале 2020 года активно обосновывала тотальный карантин, «прогнозируя» (на несколько месяцев вперед) 500 тыс. смертей в Англии и 2 млн. в США (за лето 2020 года) [Report 9].

В терминах принятых обозначений эпидемиологических моделей (см., например, [Brauer et al., 2019]) в данной работе используются различные модификации моделей SIR и SIRS, отличительной особенностью которых является разбиение инфицированных на 15 групп (подробное описание приведено в [Соколов и др., 2022]), в соответствии с длительностью заболевания (Age of Infection Model). В терминах моделей популяционной динамики и демографии используются (матричные) модели Лесли [Свирижев и др., 1978] и модели Маккендрика фон Фёрстера для популяций с возрастной структурой [Нахушев, 1995], [Эбелинг и др., 2001].

Мониторинг и построение моделей осуществлялись с марта 2021 г. Через несколько месяцев на основе анализа выявленного числа зараженных для семи популяций удалось определить (идентифицировать) некоторые биологические функции, в том числе заразность как функцию длительности болезни (рис. 2). Позднее, еще через месяц, удалось связать интенсивность выявления заболевших с количеством проведенных тестов (рис. 3).



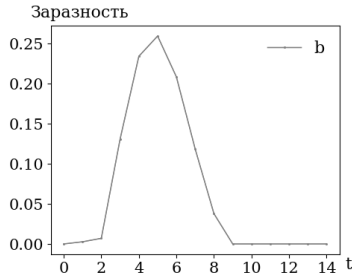


Рис. 2. Нормированная (на единицу) заразность как функция от длительности заражения (дни)

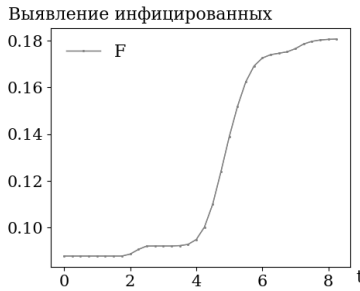


Рис. 3.  $IS\_nT$  – функция эффективности выявления больных в зависимости от количества тестов на тысячу человек ( $nT$ ) в г. Москва

Первые прогнозы были сделаны 01.12.2020 (см. рис. 4), когда удалось интегрировать в модель процесс накопления естественного (переболевших) коллективного иммунитета и его ослабление.

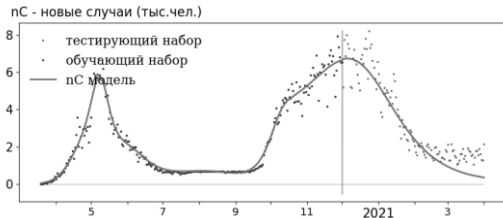


Рис. 4. Прогноз новых случаев заражения в Москве от 01.12.2020. Синие точки – обучающий набор, красная кривая – модель, зеленые точки – тестирующий набор

В апреле 2021 года при моделировании учли влияние вакцинации. Прогноз от 15.04.2021 приведен на рис. 5.

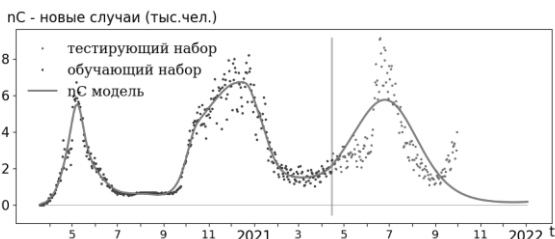


Рис. 5. Прогноз новых случаев заражения в Москве от 15.04.2021

Далее, учли ослабление иммунитета иммунизации – прогноз от 20.10.2021 (рис. 6). Большое расхождение прогноза и реальных данных объясняется появлением нового более контагиозного штамма – омикрона.

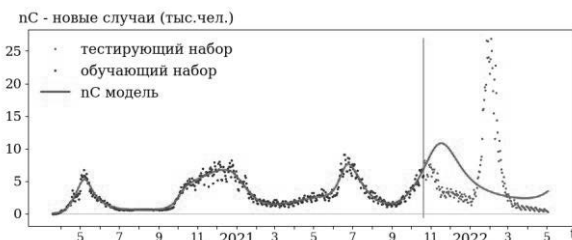


Рис. 6. Прогноз новых случаев заражения в Москве от 20.10.2021

Появления новых штаммов и их конкуренция были учтены в последней модификации модели. Соответствующий прогноз от 17.09.2022 приводится на рис. 7.

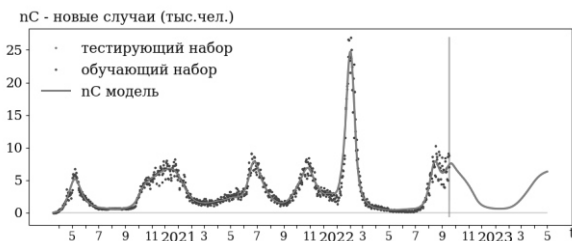


Рис. 7. Прогноз новых случаев заражения в Москве от 17.09.2022

## Заключение

В работе предложена схема построения единой многомерной классификации методов интеллектуального анализа рисков (см. рис. 1), которая разработана на основании анализа свойств социально-экономических систем, а также четырех основных способов измерения риска [Черешкин и др., 2020]. Такой подход может послужить основой для создания общей теории рисков.

В рамках инженерного подхода анализа рисков получение каких-то новых фундаментальных результатов в ближайшее время, скорее всего, проблематично. Это связано с тем, что теория вероятностей к настоящему моменту очень хорошо развита, благодаря результатам полученным, отечественными научными школами (А.Н. Колмогоров и др.). Тем не менее, если будут получены новые результаты, которые сделают работу методов, основанных на использовании нейронных сетей и машинного обучения, более прозрачными, понятными и интерпретируемыми, тут также можно обозначить некоторую перспективу.

Для модельного подхода новые результаты могут быть связаны с повсеместным внедрением новых цифровых технологий, что сделает крайне востребованным использование многоагентного подхода (например, коллективного поведения роботов, умных предприятий, дорог, городов, беспилотного транспорта и т.п.), а также различных интеллектуальных датчиков (мягкие измерения и вычисления), которым будут делегированы права самостоятельного принятия решений. В частности, авторами работы [Sokolov et al., 2020], в рамках модельного подхода анализа рисков, используется технология сбалансированной идентификации. Применение указанной технологии на практике позволило построить (на основе открытых статистических данных) модели динамики эпидемии в г. Москва с учетом 4 и 5 штаммов вируса CV19, с горизонтом прогнозирования несколько месяцев. Полученные результаты показывают, что основным механизмом, определяющим динамику, является коллективный иммунитет и его ослабление в зависимости от времени после заболевания или прививки. Предположение, что переболевшие различными штаммами получают «одинаковый» иммунитет, пока оправдывается. Проведенные численные эксперименты (см. рис. 4-7) показывают, что дальнейшее возрастание contagiозности не приводит к существенному изменению прогноза. Подробное описание математического аппарата (технология сбалансированной идентификации), использованного при построении модели представлено в работе [Sokolov et al., 2020].

В рамках экспертного подхода для задач анализа рисков упор должен быть сделан на развитие основных способов получения и представления знаний (онтологии (семантические сети), фреймы, продукции, а также

методов логического вывода (особенно с использованием неклассических логик (Финн В.К. [Финн, 2021], Вагин В. Н. [Вагин, 2019], Еремеев А.П. [Еремеев и др., 2011], Плесневич Г.С. [Плесневич и др., 2016] и др.)). Кроме того, перспективными являются подходы, ориентированные на работу с неколичественной (качественной) информацией (методы ВАР [Ларичев, 2006], нечеткая логика и мягкие вычисления [Борисов и др., 2012], работы Попова Э.В. [Попов, 2004] и т.п.).

Но наиболее существенные новые результаты могут быть получены в рамках социологического подхода анализа рисков. Прежде всего, нужно отметить работы нобелевских лауреатов по экономике Г. Саймона [Саймон, 2004], Д. Канемана [Канеман и др., 2021] и Р. Талера [Талер, 2017], связанные с исследованиями в области поведенческой экономики [Кузнецов, 2019]. В этой связи, совершенно новый импульс развития получат идеи Д. А. Поспелова (ситуационное управление [Поспелов, 2021], семиотическое моделирование, психоника (психология поведения искусственных систем) и др.) и О.И. Ларичева (психологические обоснования методов принятия решений) [Larichev, 1984]. Для развития социологического подхода анализа рисков важные результаты получены при использовании ДСМ-метода [Финн, 2021], [Михеенкова, 2009].

### Список литературы

- [Асанов и др., 2001] Асанов А.А., Ларичев О.И., Ройзензон Г.В. и др. Метод многокритериальной классификации ЦИКЛ и его применение для анализа кредитного риска // Экономика и математические методы. – 2001. – Т. 37, № 2. – С. 14-21.
- [Бек, 2000] Бек У. Общество риска. На пути к другому модерну. – М.: Прогресс-Традиция, 2000. – 384 с.
- [Борисов и др., 2012] Борисов В.В., Круглов В.В., Федулов А.С. Нечеткие модели и сети. – 2-е изд. – М.: Горячая линия–Телеком, 2012. – 284 с.
- [Бритков и др., 2018] Бритков В.Б., Ройзензон Г.В. Междисциплинарный подход к анализу риска // IX Московская международная конференция по исследованию операций (ORM2018). Труды. В 2-х т. Т. 2. / под ред. Ф.И. Ерешко. – М.: МАКС Пресс, 2018. – С. 340-345.
- [Вагин, 2019] Вагин В.Н. Знания и убеждения в интеллектуальном анализе данных. – М.: Физматлит, 2019. – 536 с.
- [Геловани и др., 2018] Геловани В.А., Бритков В.Б., Дубовский С.В. СССР и Россия в глобальной системе (1985-2030): Результаты глобального моделирования. – М.: Либроком, 2018. – 320 с.
- [Гладков и др., 2006] Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы / под ред. В.М. Курейчика. – 2-е изд. – М.: Физматлит, 2006. – 320 с.
- [Еремеев и др., 2011] Еремеев А.П., Куриленко И.Е. Темпоральные модели на основе логики ветвящегося времени в интеллектуальных системах // Искусственный интеллект и принятие решений. – 2011. – № 1. – С. 14–26.

- [**Канеман и др., 2005**] Канеман Д., Словик П., Тверски А. Принятие решений в неопределенности: правила и предубеждения. – Харьков: Гуманитарный центр, 2005. – 632 с.
- [**Канеман и др., 2021**] Канеман Д., Сибони О., Санстейн К. Р. Шум. Несовершенство человеческого суждений. – М.: Изд-во АСТ, 2021. – 544 с.
- [**Кобринский и др., 2022**] Кобринский Б.А., Зарубина Т.В., Белоносов С.С. Медицинская информатика: учебник / под ред. Т.В. Зарубиной, Б.А. Кобринского. – М.: ГЭОТАР-Медиа, 2022. – 464 с.
- [**Королев и др., 2011**] Королев В.Ю., Бенинг В.Е., Шоргин С.Я. Математические основы теории риска: учебное пособие. – 2-е изд. – М.: Физматлит, 2011. – 620 с.
- [**Кузнецов, 2019**] Кузнецов О.П. Ограниченная рациональность и принятие решений // Искусственный интеллект и принятие решений. – 2019. – № 1. – С. 3-15.
- [**Ларичев, 2006**] Ларичев О.И. Вербальный анализ решений. – М.: Наука, 2006. – 181 с.
- [**Лефевр, 2003**] Лефевр В.А. Алгебра совести. – М.: «Когито-Центр», 2003. – 426 с.
- [**Михеенкова, 2009**] Михеенкова М.А. О принципах формализованного качественного анализа социологических данных // Информационные технологии и вычислительные системы. – 2009. – № 4. – С. 40–56.
- [**Нахушев, 1995**] Нахушев А.М. Уравнения математической биологии: учебное пособие для университетов. – М.: Высшая школа, 1995. – 301 с.
- [**Плесневич и др., 2016**] Плесневич Г.С., Нгуен Т.М.В. Алгоритмы дедукции для некоторых расширений интервальной логики Аллена // Искусственный интеллект и принятие решений. – 2016. – № 1. – С. 75-88.
- [**Попков и др., 2021**] Попков Ю.С., Дубнов Ю.А., Попков А.Ю. Прогнозирование развития эпидемии COVID-19 в странах Европейского союза с использованием энтропийно-рандомизированного подхода // Информатика и автоматизация. – 2021. – Т. 20, № 5. – С. 1010-1033.
- [**Попов, 2004**] Попов Э.В. Общение с ЭВМ на естественном языке. – 2-е изд. – М.: УРСС, 2004. – 360 с.
- [**Поспелов и др., 1990**] Поспелов Д.А., Шустер В.А. Нормативное поведение в мире людей и машин. – Кишинев: Штиинца, 1990. – 136 с.
- [**Поспелов, 2021**] Поспелов Д.А. Ситуационное управление. Теория и практика. – 2-е изд. – М.: URSS, 2021. – 288 с.
- [**Ройзензон, 2005**] Ройзензон Г.В. Способы снижения размерности признакового пространства для описания сложных систем в задачах принятия решений // Новости искусственного интеллекта. – 2005. – № 1. – С. 18–28.
- [**Ройзензон, 2012**] Ройзензон Г.В. Синергетический эффект в принятии решений // Системные исследования. Методологические проблемы. Ежегодник / под ред. Ю.С. Попкова, В.Н. Садовского, В.И. Тищенко. – 2011-2012. – № 36. – М.: УРСС, 2012. – С. 248-272.
- [**Романюха, 2012**] Романюха А.А. Математические модели в иммунологии и эпидемиологии инфекционных заболеваний. – М.: Бином. Лаборатория знаний, 2012. – 293 с.

- [Саймон, 2004] Саймон Г. Науки об искусственном. – 2-е изд. – М.: Едиториал УРСС, 2004. – 144 с.
- [Свирижев и др., 1978] Свирижев Ю.М., Логофет Д.О. Устойчивость биологических сообществ. – М.: Наука, 1978. – 352 с.
- [Смирнов и др., 2022] Смирнов А.В., Молл Е.Г., Тесля Н.Н. Учет влияния человеческих факторов на процесс принятия социо-ориентированных решений при госпитализации в условиях эпидемии // Информационные технологии и вычислительные системы. – 2022. – № 3. – С. 79-94.
- [Смолян и др., 2002] Смолян Г.Л., Солнцева Г.Н. Человеческий фактор в обеспечении безопасности информационной инфраструктуры // Информационное общество. – 2002. – № 3. – С. 34-37.
- [Соколов и др., 2021] Соколов А.В., Ройзензон Г.В., Комендантова Н.П. Технология создания систем мониторинга и прогноза состояния опасных явлений и объектов (на примере эпидемии Covid-19) // Проблемы управления безопасностью сложных систем (ПУБСС-2021): Материалы XXIX международной конференции / под ред. А.О. Калашникова, В.В. Кульбы. – М.: ИПУ РАН, 2021. – С. 93-99. – Doi: 10.25728/icss.2021.46.97.011.
- [Соколов и др., 2022] Соколов А.В., Соколова Л.А. Мониторинг и прогноз динамики заболеваемости COVID-19 в Москве: 2020–2021 годы // Эпидемиология и Вакцинопрофилактика. – 2022. – Т. 21, № 4. – С. 48-59. – Doi:10.31631/2073-3046-2022-21-4-48-59.
- [Талев, 2018] Талев Н.Н. Одураченные случайностью. О скрытой роли шанса в бизнесе и в жизни. – 5-е изд. – М.: Манн, Иванов и Фербер, 2018. – 313 с.
- [Талер, 2017] Талер Р. Новая поведенческая экономика. Почему люди нарушают правила традиционной экономики. – М.: Эксмо, 2017. – 368 с.
- [Финн, 2021] Финн В.К. Применение интеллектуальных систем в изучении общества: Методы логики и искусственного интеллекта в гуманитарном знании, строение интеллектуальных систем и социальные последствия их использования. – М.: URSS, 2021. – 352 с.
- [Черешкин и др., 2020] Черешкин Д.С., Ройзензон Г.В., Бритков В.Б. Применение методов искусственного интеллекта для анализа риска в социально-экономических системах // Информационное общество. – 2020. – № 3. – С. 14–24.
- [Эбелинг и др., 2001] Эбелинг В., Энгель А., Файстель Р. Физика процессов эволюции. – М.: Едиториал УРСС, 2001. – 328 с.
- [Brauer et al., 2019] Brauer F., Castillo-Chavez C., Feng Z. *Mathematical Models in Epidemiology*. – New York: Springer, 2019. – 637 pp.
- [Cook et al., 1982] Cook R.D., Weisberg S. Criticism and Influence Analysis in Regression // *Sociological Methodology*. – 1982. – Vol. 13. – P. 313-336.
- [Kermack et al., 1927] Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics // *Proceedings of the Royal Society of London*. – 1927. – Vol. 115, No. 772. – P. 700-721.
- [Larichev, 1984] Larichev O.I. Psychological Validation of Decision Methods // *Journal of Applied Systems Analysis*. – 1984. – Vol. 11. – P. 37-46.

- [Furems et al., 2003] Furems E.M., Larichev O.I., Roizenon G.V., et al. Human behavior in a multi-criteria choice problem with individual tasks of different difficulties // International Journal of Information Technology and Decision Making. – 2003. – Vol. 2, No. 1. – P. 29-40.
- [Salgotra et al., 2021] Salgotra R., Gandomi A. H. Time series analysis of the COVID-19 pandemic in Australia using genetic programming. Data Science for COVID-19. – 2021. – P. 399-411. – Doi: 10.1016/B978-0-12-824536-1.00036-8.
- [Sokolov et al., 2020] Sokolov A.V., Voloshinov V.V. Model selection by balanced identification: the interplay of optimization and distributed computing // Open Computer Science. – 2020. – Vol. 10, No. 1. – P. 283-295.
- [Rahimi et al., 2021] Rahimi I, Chen F, Gandomi AH. A review on COVID-19 forecasting models // Neural Comput Appl. – 2021. – Feb 4:1-11. – Doi: 10.1007/s00521-020-05626-8.
- [Report 9] Электрон. Ресурс. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare. – URL: <https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-16-COVID19-Report-9.pdf> (дата обращения: 17.09.2022)
- [Tamang, et al., 2020] Tamang S.K., Singh P.D., Datta, B. Forecasting of Covid-19 cases based on prediction using artificial neural network curve fitting technique // Global Journal of Environmental Science and Management. – 2020. – No. 6. – P. 53-64. – Doi: 10.22034/GJESM.2019.06.SI.06.

## ОЦЕНКА РИСКА ВНЕЗАПНОЙ СЕРДЕЧНОЙ СМЕРТИ С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ\*

Н.И. Юсупова (*yusupova.ni@ugatu.su*)  
М.Р. Богданов (*bogdanov\_marat@mail.ru*)  
Уфимский государственный авиационный  
технический университет, Уфа

Работа посвящена обнаружению предиктора внезапной сердечной смерти – фибрилляции предсердий. Использовались общедоступные базы данных оцифрованных электрокардиограмм. Было использовано два подхода: машинное обучение и цифровые биомаркеры. Была установлено, что с помощью метода Category boost можно выявлять фибрилляцию предсердий на фоне других форм аритмии и нормального синусового ритма с эффективностью 99%. В качестве цифровых биомаркеров были использованы ортогональные вейвлеты, полученные из двухсекундных образцов ECG Fragment Database for the Exploration of Dangerous Arrhythmia. Была показана высокая селективность вейвлетов к фибрилляции предсердий.

**Ключевые слова:** предикторы внезапной сердечной смерти, машинное обучение, вейвлет-анализ.

### Введение

Согласно данным Всемирной организации здравоохранения заболевания сердечно-сосудистой системы являются одной из основных причин смертности. Остановка сердца может быть вызвана разными причинами, в том числе устойчивой желудочковой тахикардией, включая желудочковую тахикардию (мономорфную или полиморфную), трепетанием желу-

---

\* Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках основной части государственного задания высшим учебным заведениям No FEUE-2020-0007, РНФ 22-19-00471 Система поддержки принятия решений для профилактики и лечения бронхолегочных заболеваний, оценки рисков заболеваний и осложнений их лечения в задачах персонализированной медицины на основе методов анализа данных и искусственного интеллекта.



дочков или фибрилляцией желудочков [1]. Особую опасность представляет так называемая внезапная сердечная смерть (ВСС), вызванная очень частыми неритмичными сокращениями желудочков сердца. ВСС характеризуется как смерть, произошедшая при свидетеле в течение часа с момента появления новых симптомов или смерть без свидетелей в промежутке от 1 до 24 часов без какой-либо известной причины смерти. При этом кардиомиоциты начинают сокращаться хаотично, разрозненно, и сердце перестает работать как насос, подающий кровь к другим органам и тканям. Синдром внезапной сердечной смерти может быть обусловлен широким спектром различных электрических и механических факторов, в том числе острым инфарктом миокарда, хронической коронарной болезнью с предшествующим инфарктом(ами) миокарда, кардиомиопатиями, миокардитом, пороками клапанов сердца, дисплазией правого желудочка, синдромом удлиненного интервала Q-T, преждевременным возбуждением Вольфа-Паркинсона-Уайта, синдромом Бругада, кардиоактивными препаратами такими, как адреналин, кокаин и др. [2]. У некоторых людей нет очевидной электрической или механической предрасположенности к внезапной сердечной смерти.

Фибрилляция предсердий (ФП) является наиболее распространенной формой аритмии, от которой страдает свыше 33 миллионов человек во всем мире. ФП связана с повышенным риском инсульта и сердечной недостаточности (СН). Недавние исследования говорят о том, что ФП также связана с повышенным риском внезапной сердечной смерти (ВСС) [3, 4]. Важность раннего обнаружения предикторов ВСС подтверждает тот факт, что в недавно появившихся на рынке портативных устройствах с функцией регистрации ЭКГ реализуются алгоритмы, выявляющие опасные нарушения ритма. Так, в смарт-часах компании Apple реализована функция обнаружения нормального синусового ритма и фибрилляции предсердий [5]. Портативный электрокардиограф Alivecor kardia позволяет выявлять фибрилляцию предсердий и удлинение QT-интервала, приборы Omron heartscan и Zenicor ECG также позволяют выявлять ФП [6].

## **Цели и задачи**

Целью предлагаемого исследования является разработка алгоритмов выявления предикторов внезапной сердечной смерти.

## **Литературный обзор**

Мерцательная аритмия или фибрилляция предсердий (ФП) является наиболее распространенной аритмией, которой страдает около 1% населения. При этом 5% лиц в возрасте 65 лет и старше имеют ФП [3, 4]. Кардиологи советуют проводить скрининг на ФП у пациентов, находящихся в

группе риска независимо от возраста, и проводить скрининг всех пациентов в возрасте 65+ лет [7]. ФП характеризуется учащенным, нерегулярным сердцебиением [8]. Распространенными причинами болезни являются: артериальная гипертензия, кардиомиопатия, ишемия и ревматическая болезнь сердца [7]. Нерегулярное сердцебиение может вызвать образование тромбов, что увеличивает риск инсульта в пять раз [1] и риск смерти в два раза по сравнению с людьми того же возраста, независимо от других факторов риска [1]. Патофизиологические механизмы ФП до конца не изучены, что затрудняет диагностику [1]. По некоторым оценкам, 20% всех случаев ФП могут остаться незамеченными, поскольку заболевание протекает бессимптомно. При таких обстоятельствах лечение не начинают, что может привести к негативным последствиям для пациентов. Таким образом, разработка новой технологии обнаружения ФП может помочь в скрининге и лечении, снижая заболеваемость и смертность.

Наиболее предпочтительным методом выявления ФП является двенадцатиканальная электрокардиография. На рис. 1 сравниваются нормальный синусовый ритм (слева) и фибрилляция предсердий. В случае патологии наблюдаются неравные промежутки между R-зубцами, а зубец P приглушен или отсутствует.

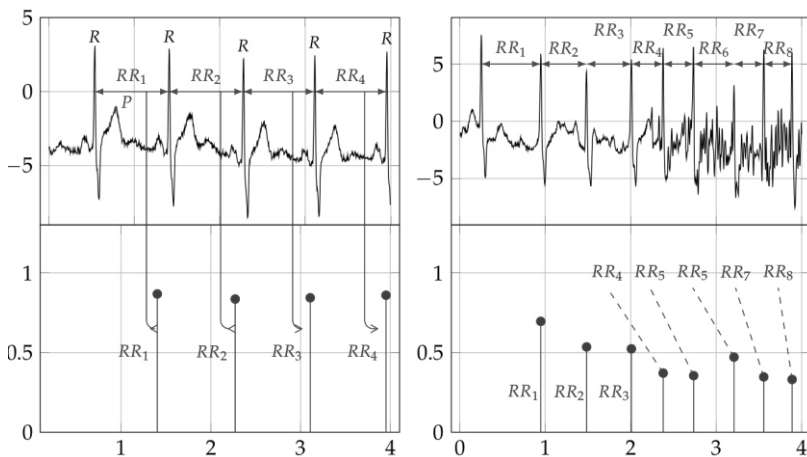


Рис. 1. Сравнение нормального синусового ритма (слева) и фибрилляции предсердий (справа)

В верхней части рисунка показаны электрокардиограммы (по оси ординат отложена амплитуда в милливольтгах), в нижней части рисунка – расстояния между R-зубцами (по оси ординат отложено время в секундах). По осям абсцисс отложено время в секундах

Для автоматизированного обнаружения ФП можно использовать два подхода. В первом случае из оцифрованных электрокардиограмм выделяются признаки, поступающие на вход методов машинного обучения. Во втором случае из оцифрованных электрокардиограмм получают цифровые биомаркеры, которые в последствие также поступают на вход методов машинного обучения. При этом пациент выступает в качестве источника сигналов, а врач – получателя.

Из работ, посвященных автоматическому распознаванию оцифрованных электрокардиограмм можно выделить статьи Пана и Томкинса [6] и Мохаммеда Качуи [7].

При классификации признаков, выделенных из кардиосигнала используются различные методы машинного обучения, например, методы обучения без учителя (k-ближайших соседей, скрытые марковские модели и генеративно-дискриптивные нейронные сети [8]) и с учителем (наивный байес (NB), вероятностная нейронная сеть (PNN), метод опорных векторов (SVM), случайный лес (RF), нейронная сеть Левенберга-Марквардта (LMNN), дерево решений (DT), основанное на правилах и случайный лес (RF).

Обычно авторы исследований в области автоматического обнаружения ФП использовали общедоступные базы данных оцифрованных электрокардиограмм, такие как MIT-BIH AFIB, MIT-BIH Arrhythmia Database, MIT-BIH NSR. Среди работ, в которых использовались цифровые биомаркеры, можно отметить исследования Wang et al., 2020 [9] (вейвлет-пакетная декомпозиция с последующими многомерными статистическими функциями), Cao et al., 2020 [10] (аугментация данных), Marsili et al., 2019 [8] (Шенноновская энтропия), Yao et al., 2019 [8] (цифровое вейвлет-преобразование), Lui et al., 2018 [8] (нормализованная размытая энтропия), Xia et al., 2018 [11] (синхронное сжатие кратковременного преобразования Фурье), Majia et al., 2013 [8] (разложение по эмпирическим модам (emd) и статистика высшего порядка (hos). Широкое распространение в подобных исследованиях получили такие методы машинного обучения, как метод опорных векторов (Asgari et al., 2015 [12]), сверточная нейронная сеть (Anderson et al., 2019 [8]), случайный лес (Kennedy et al., 2016 [13]). В ряде исследований ФП выявляется на основе анализа RR-интервалов. В качестве примеров можно привести работы Ivanovic et al., 2019 [14], Anderson et al., 2019 [8] и Faust et al., 2018 [15].

## **Материалы и методы**

В рамках проводимых исследований использовались общедоступные базы данных оцифрованных электрокардиограмм, размещенных на ресурсе [physionet.org](http://physionet.org).

1. База данных фрагментов электрокардиограмм для изучения опасных форм аритмии (ECG Fragment Database for the Exploration of Dangerous Arrhythmia, <https://physionet.org/content/ecg-fragment-high-risk-label/1.0.0/>) [16, 17].

База данных содержит набор из 2-х секундных фрагментов сигналов ЭКГ с нарушениями ритма с частотой дискретизации 180 Гц, два отведения, которые сгруппированы в отдельные классы по степени угрозы для жизни пациента, в том числе:

1. Жизнеугрожающие аритмии, требующие неотложных реанимационных мероприятий:

- VFL: трепетание желудочков;
  - ФЖ: фибрилляция желудочков.
2. Особая форма жизнеугрожающих аритмий:
- VTTdP: желудочковая тахикардия torsade de pointes.
3. Жизнеугрожающие желудочковые аритмии:
- VTHR: желудочковая тахикардия высокой частоты (мономорфная и полиморфная).
4. Потенциально опасные желудочковые аритмии:
- VTLR: желудочковая тахикардия низкой частоты (мономорфная и полиморфная);
  - В: желудочковая бигеминия;
  - HGEA: высокая степень желудочковой эктопической активности;
  - VER: выскальзывающий желудочковый ритм.
5. Наджелудочковые аритмии:
- AFIB: фибрилляция предсердий;
  - СВТА: наджелудочковая тахикардия;
  - SBR: синусовая брадикардия;
  - ВI: блокада сердца первой степени;
  - NOD: узловой (av) ритм.
6. Синусовый ритм:
- BBB: синусовый ритм с блокадой ножки пучка Гиса;
  - N: нормальный синусовый ритм;
  - Ne: нормальный ритм с одиночной экстрасистолой.

В качестве основы была выбрана База данных злокачественных желудочковых эктопий MIT-BIH (MVED).

2. База данных электрокардиограмм с нормальным синусовым ритмом (MIT-BIH Normal Sinus Rhythm Database, <https://physionet.org/content/nsrdb/1.0.0/>) [18]. Частота дискретизации – 128 Гц, 2 отведения. Эта база данных включает 18 длительных записей ЭКГ субъектов, направленных в

лабораторию аритмии бостонской клиники Бет Исраэль (ныне Медицинский центр имени дьяконессы Бет Исраэль). У испытуемых, включенных в эту базу данных, не было значительных аритмий; среди них 5 мужчин в возрасте от 26 до 45 лет и 13 женщин в возрасте от 20 до 50 лет.

3. База данных аритмий (**MIT-BIH Arrhythmia Database**, <https://physionet.org/content/mitdb/1.0.0/>). [19]. База данных содержит 48 получасовых фрагментов двухканальных записей ЭКГ, полученных от 47 субъектов. Частота дискретизации – 360 Гц, разрядность – 11 бит. Два или более кардиологов независимо аннотировали каждую запись (всего около 110 000 аннотаций)

4. База данных фибрилляции предсердий (**MIT-BIH Atrial Fibrillation Database**, <https://physionet.org/content/afdb/1.0.0/>). [20]. Включает в себя 25 длительных записей ЭКГ испытуемых с мерцательной аритмией (в основном пароксизмальной). Каждая из записей имеет продолжительность 10 часов и содержит два отведения, частота дискретизации – 250 Гц, разрядность – 12 бит, разрешение в диапазоне  $\pm 10$  мВ.

5. База данных наджелудочных аритмий (**MIT-BIH Supraventricular Arrhythmia Database**, <https://physionet.org/content/svdb/1.0.0/>). [21]. БД содержит 78 получасовых записей с частотой дискретизации 128 Гц в двух отведениях.

6. База данных злокачественных желудочковых эктопий MIT-BIH (**MIT-BIH Malignant Ventricular Ectopy Database**, <https://physionet.org/content/vfdb/1.0.0/>) [22–24]. Эта база данных включает 22 получасовых записи ЭКГ субъектов, у которых были эпизоды устойчивой желудочковой тахикардии, трепетания желудочков и фибрилляции желудочков. Частота дискретизации – 250 Гц, 2 отведения.

В рамках проводимых исследований основное внимание было посвящено обнаружению фибрилляции предсердий. Для этого были использованы два подхода. В первом случае задача решалась в рамках классической концепции машинного обучения. Во втором случае использовались цифровые биомаркеры фибрилляции предсердий.

Было разработано программное обеспечение, написанное на языке Python. Использовалась вычислительная система со следующими характеристиками: Intel Core i9, 128 Gb RAM, NVIDIA RTX 3080 Ti.

## Машинное обучение

Как уже говорилось выше, частота дискретизации исходных кардиосигналов варьирует в пределах от 128 до 360 Гц. Было принято решение взять за основу наименьшую частоту дискретизации. Препроцессинг проводили следующим образом. В начале из всех кардиосигналов были выделены QRST-комплексы, их частота дискретизации была снижена до 128 Гц. В результате были получены векторы признаков длиной 76 элементов (рис. 2).

## *Рис. 2. Препроцессинг кардиосигналов*

Затем была проведена бинарная классификация с помощью метода Category Boost, разработанного специалистами компании Yandex для прогнозной аналитики и рекомендательных систем. (<https://catboost.ai/en/docs/>). Выбор метода классификации был обусловлен результатами предварительных исследований, в ходе которых сравнивалась эффективность распознавания как метод опорных фекторов, полносвязная нейронная сеть, сверточная нейронная сеть и Category Boost. Последний метод показал лучшую метрику ассурасу.

### **Использование цифровых биомаркеров**

В качестве цифровых биомаркеров были использованы ортогональные вейвлеты, полученные из двухсекундных образцов ECG Fragment Database for the Exploration of Dangerous Arrhythmia.

Вейвлет – это некая функция, заданная таблично (небольшой массив чисел), определенным образом характеризующая сигнал. Среди этих функций важную роль играют ортогональные вейвлеты. У такого вейвлета коэффициент корреляции с целевым сигналом стремится к единице, в то время как по отношению к другим сигналам корреляция этого вейвлета стремится к нулю. Для создания вейвлета необходимо избавиться от избыточности исходного сигнала и выявить характерную область, адекватно описывающую этот сигнал. Устранение избыточности сигнала производится с помощью процедуры детектирования.

Степень разреженности исходного сигнала задается размером так называемой рамки детектирования. Выбранная область детектированного сигнала подвергается процедуре нормирования. Нормированный участок

детектированного сигнала и является вейвлетом. Нормирование позволяет устранить влияние амплитуды сигнала на процесс распознавания. Важно определить оптимальную длину вейвлета. Длинные вейвлеты более информативны, но могут давать ошибочную корреляцию с нецелевыми сигналами. Создав хорошие (ортогональные) вейвлеты, мы можем использовать их в дальнейшем для распознавания неизвестных сигналов.

В ходе экспериментов было установлено, что оптимальной длиной вейвлета является 70 отсчетов, оптимальный размер рамки детектирования – 20 отсчетов. Рассмотрим процедуру создания вейвлета на примере двухсекундных образцов аритмий и электрокардиограмм синусового ритма. Вначале необходимо считать исходный сигнал и записать его в массив (рис. 3).

*Рис. 3. Исходный кардиосигнал*

После этого требуется избавиться от избыточности сигнала. Для этого осуществляется процедура детектирования сигнала. Предположим, исходный сигнал записан в массив  $A_i$ ,  $i$  изменяется от 1 до  $N$ . Задается рамка считывания сигнала. Малый размер рамки повышает информативность, но оборачивается большими накладными расходами в плане производительности. Как уже говорилось выше, размер рамки был определен в 20 единиц. После этого весь диапазон массива  $A$  делится на  $N/20$  рамок. Внутри каждой рамки находили сумму квадратов  $A_i$ . Далее, получаем массив детектированного сигнала  $B_j$ , где элементы массива  $B$  вычисляются как корень квадратный из  $sum_j/20$ . На рис. 4 показаны детектированные сигналы.

*Рис. 4. Детектированный электрокардиосигнал*

Произведя детектирование сигнала, выбираем область, характерную для конкретного сигнала. Мы остановились на длине вейвлета в 70 отсчетов. Выбрав характерную область детектированного сигнала  $V_j$ , где  $j$  изменяется от  $V_{min}$  до  $V_{max}$ , осуществляем нормирование детектированного сигнала  $V_j$  на отрезке от  $V_{min}$  до  $V_{max}$ . Нормированный детектированный сигнал и является вейвлетом (рис. 5).

*Рис. 5. Вейвлет электрокардиосигнала*

Процесс распознавания проходил следующим образом. На вход системы подается неизвестный сигнал  $U_i$ . Далее, производится детектирование сигнала с такой же рамкой (в нашем случае 20 отсчетов), что и при создании вейвлета. Получаем массив детектированного сигнала  $C_j$  (в 20 раз меньше отсчетов, чем в исходном сигнале). Далее, выбирается шаг распознавания  $Step$ , равный размеру вейвлета  $m$  (в нашем случае  $m=70$ ). Внутри шага распознавания определяется коэффициент корреляции участка ис-



следуемого сигнала  $C_j$ , равного  $Step$ , со сравниваемым вейвлетом. Мы использовали коэффициент корреляции Пирсона. Найдя коэффициент корреляции на выбранном шаге, сдвигаем рамку распознавания на один элемент и возобновляем процесс (рис. 6).

*Рис. 6. Коэффициент корреляции между вейвлетом 2-х секундного фрагмента аритмии и детектированным сигналом синусового ритма*

Если коэффициент корреляции неизвестного сигнала и известного вейвлета оказывается близок к единице хотя бы раз, программа принимает решение об успехе распознавания.

## Результаты

Результаты бинарной классификации признаков, выделенных из различных кардиосигналов, приведены в табл. 1. В качестве метода классификации использовался Category Boost.

Таблица 1

### Результаты бинарной классификации

1-я выборка	2-я выборка	Accuracy
mit_bih_arrythmia	<b>atrial_fibrillation</b>	0.99
normal sinus rythm	<b>atrial_fibrillation</b>	0.99
normal sinus rhythm	arrhythmia	0.99
supraventricular_arrhythmia	<b>atrial_fibrillation</b>	0.99
mit_bih_malignant_ventricular_ectopy	<b>atrial_fibrillation</b>	0.99

Из табл. 1 видно, что фибрилляция предсердий хорошо распознается в случае попарного сравнения с синусовым ритмом, а также при сравнении с другими формами аритмий.

На следующем этапе исследований были сформированы 6 выборок ортогональных вейвлетов, характерных для различных форм аритмий и синусового ритма. В табл. 2 приведены результаты сканирования таких вейвлетов вдоль различных кардиосигналов. Числовые значения – коэффициента корреляции.

В литературе описаны примеры использования различных вейвлетов в качестве признаков, используемых при классификации кардиологических заболеваний, в частности, описано использование вейвлетов Добеши. Использование ортогональных вейвлетов при распознавании ЭКГ в литературе нам не встречалось.

Из табл. 2 видно, что вейвлеты хорошо настроены на фибрилляцию предсердий. Предикторы опасных нарушений сердечно-сосудистой системы, таких как фибрилляция предсердий могут проявляться очень редко. Для повышения эффективности алгоритмов распознавания ФП желательно увеличивать время регистрации ЭКГ. Эту задачу можно решать с помощью онлайн-мониторинга пациентов с использованием технологии интернета медицинских вещей.

Таблица 2

### Оценка эффективности вейвлет-анализа

	sinus-rhythm	supraventricular arrhythmia	atrial fibrillation	malignant-ventricular-ectopy	arrhythmia
Dangerous_VFL_VF	0,67	0,70	<b>0,91</b>	0,47	0,44
Potential_Dangerous	0,69	0,66	<b>0,92</b>	0,47	0,45
Sinus_rhythm	0,77	0,73	<b>0,95</b>	0,52	0,47
Special_Form_VTTdP	0,67	0,68	<b>0,90</b>	0,46	0,46
Supraventricular	0,68	0,69	<b>0,96</b>	0,46	0,48
Threatening_VT	0,72	0,70	<b>0,93</b>	0,48	0,44

В качестве практической реализации предложенных методов распознавания ЭКГ можно предложить следующие сценарии:

1. Онлайн мониторинг состояния пациентов. С помощью вейвлетов анализируются двухсекундные временные интервалы ЭКГ с целью выявления предикторов внезапной сердечной смерти.

2. Постобработка результатов холтеровского мониторирования. Длительные записи ЭКГ анализируются с помощью метода Category Boost.

## Список литературы

- Greenwald, Scott David.** The development and analysis of a ventricular fibrillation detector. Thesis (M.S.) – Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 1986.
- Greenwald S.D., Albrecht P., Moody G.B., Mark R.G.** Estimating confidence limits for arrhythmia detector performance // *Computers in Cardiology*. – 1985. – 12. – P. 383-386.
- Victor Waldmann, Xavier Jouven, Kumar Narayanan, Olivier Piot, Sumeet S. Chugh, Christine M. Albert and Eloi Marijon.** Association Between Atrial Fibrillation and Sudden Cardiac Death. Pathophysiological and Epidemiological Insights. Originally published 2 Jul 2020. – <https://doi.org/10.1161/CIRCRESAHA.120.316756> *Circulation Research*. – 2020. – 127. – P. 301-309. – <https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.120.316756>.
- Chen L.Y., Benditt D.G., Alonso A.** Atrial fibrillation and its association with sudden cardiac death // *Circ J*. – 2014. – 78 (11). – P. 2588-93. – Doi: 10.1253/circj.cj-14-0814. Epub 2014 Sep 26. PMID: 25262841.  
<https://support.apple.com/en-us/HT208955>.
- Pan J., Tompkins W.J.** A real-time QRS detection algorithm // *IEEE Trans Biomed Eng*. – 1985 Mar. – 32 (3). – 230-6. – Doi: 10.1109/TBME.1985.325532. PMID: 3997178.
- Mohammad Kachuee, Shayan Fazeli, Majid Sarrafzadeh.** ECG Heartbeat Classification: A Deep Transferable Representation // *arXiv:1805.00794v2 [cs.CY]* 12 Jul 2018.
- Agam Bansal, Rajnish Joshi.** Portable out-of-hospital electrocardiography: A review of current technologies // *Journal of Arrhythmia*. – 2018. – 34. – P. 129-138.
- Moody G.B., Mark R.G.** The impact of the MIT-BIH Arrhythmia Database // *IEEE Eng in Med and Biol*. – 20 (3). – P. 45-50 (May-June 2001). (PMID: 11446209).
- Oliver Faust, Edward J. Ciaccio and U. Rajendra Acharya.** A Review of Atrial Fibrillation Detection Methods as a Service // *Int. J. Environ. Res. Public Health*. – 2020. – 17. – P. 3093. – Doi:10.3390/ijerph17093093. [www.mdpi.com/journal/ijerph](http://www.mdpi.com/journal/ijerph).
- Wang, J., Wang, P., Wang, S.** Automated detection of atrial fibrillation in ECG signals based on wavelet packet transform and correlation function of random process // *Biomed. Signal Process. Control*. – 2020. – 55. – P. 101662.
- Cao P., Li X., Mao K., Lu F., Ning G., Fang L., Pan Q.** A novel data augmentation method to enhance deep neural networks for detection of atrial fibrillation // *Biomed. Signal Process. Control*. – 2020. – 56. – P. 101675.
- Xia Y., Wulan N., Wang K., Zhang H.** Detecting atrial fibrillation by deep convolutional neural networks // *Comput. Biol. Med*. – 2018. – 93. – P. 84-92.
- Asgari S., Mehrnia A., Moussavi M.** Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine // *Comput. Biol. Med*. – 2015. – 60. – P. 132-142.
- Kennedy A., Finlay D.D., Guldenring D., Bond, R.R., Moran K., McLaughlin J.** Automated detection of atrial fibrillation using RR intervals and multivariate-based classification // *J. Electrocardiol*. – 2016. – 49. – P. 871-876.

- Ivanovic M.D., Atanasoski V., Shvilkin A., Hadzиеvski L., Maluckov A.** Deep Learning Approach for Highly Specific Atrial Fibrillation and Flutter Detection based on RR Intervals // In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019. – P. 1780-1783.
- Faust O., Shetty R., Sree S.V., Acharya S., Acharya R., Ng E., Poo C.K., Suri J.** Towards the systematic development of medical networking technology // *J. Med. Syst.* – 2011. – 35. – P. 1431-1445.
- Nemirko A., Manilo L., Tatarinova A., Alekseev B., & Evdakova E.** ECG Fragment Database for the Exploration of Dangerous Arrhythmia (version 1.0.0). PhysioNet. – <https://doi.org/10.13026/kpfg-xs25>, 2022.
- Manilo L.A., Nemirko A.P., Evdakova E.G. and Tatarinova A.A.** ECG Database for Evaluating the Efficiency of Recognizing Dangerous Arrhythmias // 2021 IEEE Ural-Siberian Conference on Computational Technologies in Cognitive Science, Genomics and Biomedicine (CSGB). – 2021. – P. 120-123. – Doi: 10.1109/CSGB53040.2021.9496029.
- Goldberger A., Amaral L., Glass L., Hausdorff J., Ivanov P.C., Mark R., ... & Stanley H. E.** PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*. – 2000. – 101 (23). – P. e215-e220.
- Moody G.B., Mark R.G.** The impact of the MIT-BIH Arrhythmia Database // *IEEE Eng. in Med. and Biol.* – 20 (3). – P. 45-50 (May-June 2001). (PMID: 11446209).
- Moody G.B., Mark R.G.** A new method for detecting atrial fibrillation using R-R intervals. *Computers in Cardiology.* – 1983. – 10. – P. 227-230.
- Greenwald S.D.** Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms using contextual information: Ph.D. thesis, Harvard-MIT Division of Health Sciences and Technology, 1990.
- Greenwald S.D.** Development and analysis of a ventricular fibrillation detector. M.S. thesis, MIT Dept. of Electrical Engineering and Computer Science, 1986.

## ИССЛЕДОВАНИЕ ПЕРЕОБУЧЕНИЯ В АЛГЕБРАИЧЕСКОМ МАШИННОМ ОБУЧЕНИИ

Л.А. Якимова (*jl10dmila@yandex.ru*)  
ФИЦ Информатика и управление РАН, Москва

В статье описываются результаты экспериментального исследования поведения решателей, основанных на бинарной операции схождения. В качестве интересующих нас феноменов исследуются классическая проблема переобучения (особенно, ее устойчивость при увеличении минимального числа порождающих примеров) и приводящее к нему понятие фантомных сходов в ДСМ- и ВКФ-системах.

**Ключевые слова:** интеллектуальный анализ данных, машинное обучение, переобучение, фантомное схождение.

### Введение

В настоящее время ежедневно появляется необходимость решения задач анализа, логистики, прогноза, выявления скрытых закономерностей в различных сферах жизнедеятельности человека. Из-за стремительного роста объемов информации в XXI веке, развития способов и технологий ее сбора и хранения в базах данных существующим методам анализа информации и моделирования изучаемых объектов неизбежно требуется автоматизация поддержки эксперта посредством интеллектуального анализа данных, классификации и машинного обучения. Вследствие этого, тема оптимизации вычислительных методов машинного обучения и сокращения времени их выполнения в различных системах, несомненно, важна на сегодняшний день, как никогда ранее.

Данная статья посвящена изучению и исследованию феномена «переобучения» и поиску фантомных гипотез в ДСМ- и ВКФ-методах. Однако стоит отметить, что феномен переобучения в других парадигмах давно известен и неоднократно изучался. Будут представлены результаты исследования устойчивости фантомных гипотез при последовательном увеличении числа их родителей, которым ранее никто не занимался.

## 1. Экспериментальное обнаружение переобучения ДСМ-решателя

Начиная с конца 70-ых годов прошлого века и по сегодняшний день, группа исследователей под руководством проф. В.К. Финна развивают так называемый ДСМ-метод автоматического порождения гипотез [Финн и др., 2009].

ДСМ-метод назван так в честь известного английского философа, экономиста и логика Джона Стьюарта Милля. В.К. Финну с коллегами удалось развить идеи индуктивной логики Милля до системы поддержки научных исследований. Основой этого подхода является использование бинарной операции сходства для обобщения обучающих примеров в гипотезы о причинах наличия/отсутствия целевого свойства. Стоит отметить, что приблизительно в это же самое время (конец 70-х годов XX века) упрощенный подход к анализу данных, основанный на теории решеток, был разработан командой зарубежных исследователей под руководством проф. Рудольфа Вилле под названием анализ формальных понятий (АФП) [Ganter et al., 1999]. Однако в отечественный подход были включены рассмотрение контр-примеров, а также абдукция и доопределение по аналогии, чего не имелось в АФП.

Благодаря трудам проф. В.К. Финна и его единомышленников был создан ряд компьютерных систем, получивших общее название ДСМ-систем интеллектуального анализа данных.

Однако имеются некоторые сложности в применении ДСМ-метода к анализу данных:

Во-первых, множество порождаемых ДСМ-гипотез может оказаться экспоненциально велико по сравнению с размером обучающей выборки.

Во-вторых, несколькими учеными были доказаны пессимистические оценки сложности для многих ДСМ-процедур, т.е. мы имеем в виду #P- и NP-полноту связанных с ними задач распознавания и перечисления.

В-третьих, была обнаружена еще одна трудность: существование так называемых «фантомных» ДСМ-гипотез, описанных в [Виноградов, 2015]. Этот эффект соответствует феномену «переобучения», о котором более детально будет рассказано далее.

### 1.1. Феномен переобучения

Переобучение в машинном обучении – нежелательное явление, при котором алгоритм обучения точно объясняет примеры из обучающей выборки, однако недостаточно хорошо предсказывает примеры из тестовой выборки.

Мы исследовали феномен переобучения для систем поддержки научных исследований, основанных на операции сходства, в которых он проявляется в том, что в части отрицательных тестовых примерах содержатся сходства положительных обучающих примеров. Это связано с тем, что

сходства положительных обучающих примеров должны соответствовать причине, вызывающей их положительное целевое свойство. Поэтому включение этого сходства в любой отрицательный тестовый пример приводит к ошибочному предсказанию. Это говорит о том, что мы наблюдаем ситуацию, когда модель, хорошо ведущая себя на обучающей выборке, делает ошибочные прогнозы на тестовой выборке.

Это связано с тем, что пытаясь максимально учесть всю имеющуюся информацию из обучающей выборки при построении модели («в процессе обучения»), обнаруживаются некоторые фантомные сходства. И, как следствие, мы получаем модель, которая имеет плохую предсказательную способность.

## 1.2. Устойчивость переобучения

Тестовые примеры доопределяются положительно, если в них вкладывается хотя бы одно сходство положительных обучающих объектов. Сходства отрицательных примеров мы не вычисляем. Тестовые примеры доопределяются отрицательно, если в них не вкладывается ни одно сходство положительных обучающих объектов. То, что во всех экспериментах все положительные тестовые примеры доопределяются правильно, означает, что сходства очень устойчивы, то есть, например, даже из половины всех примеров могут восстановиться все.

Ложное предсказание некоторого отрицательного примера, означает, что в него вкладывается некоторое сходство, для которого этот пример, если бы его поместили не в тестовую выборку, а в обучающую, стал контр-примером. Другими словами, в него вложилось «фантомное» сходство.

Для устранения «фантомных» сходств мы будем увеличивать число родителей, т.е. число обучающих примеров, в каждом из которых встречается соответствующий фрагмент описания.

В таблице ниже представлено, насколько устойчивым является феномен переобучения. Следует пояснить, что в данном случае мы последовательно увеличивали нижний порог родителей до 100–120.

	Обучающая выборка	Тестовая выборка	Положительное свойство	Результат
1	MUSHROOMS.train	MUSHROOMS.rest	быть съедобным	2120 положительных и 1972 отрицательных
2	MUSHROOMS.train	MUSHROOMS.rest	быть ядовитым	1961 из 1972 (+) и 2115 из 2120(-)
3	MUSHROOMS.rest	MUSHROOMS.train	быть съедобным	2088 (+) и 1944 (-)
4	MUSHROOMS.rest	MUSHROOMS.train	быть ядовитым	1936 из 1944 (+) 2077 из 2088 (-)

В первом и третьем случаях мы добились устранения переобучения. За счет повышения порога родителей (сохранения только устойчивых родителей) удалось полностью устранить эффект переобучения. Т.е. в этих случаях все положительные и все отрицательные примеры предсказаны правильно. В то же время во втором и четвертом случаях мы не смогли избавиться от переобучения до тех пор, пока не проявился эффект недообучения. Например, в последнем (четвертом) случае из 1944 положительных примеров правильно предсказано 1936, а из 2088 отрицательных примеров – 2077. Таким образом, в этом случае невозможно использовать исчерпывающее порождение гипотез (даже с фильтром устойчивости), чтобы одновременно полностью устранить эффекты переобучения и недообучения.

## 2. Экспериментальное обнаружение фантомных гипотез

Фантомными называют сходства, возникающие лишь из-за одновременного наличия общего фрагмента в нескольких объектах-родителях, каждый из которых имеет свою собственную фрагмент-причину.

Появление таких несущественных сходств, возникающих из-за случайных совпадений множества атрибутов у нескольких обучающих примеров, каждый из которых имеет различные реальные причины, ведет к «экспоненциальному взрыву» числа гипотез. Для того чтобы решить этот вопрос ДСМ-метод предлагает два стандартных механизма: запрет гипотез, имеющих контр-примеры, и порождение гипотез из  $\geq b$  примеров-родителей.

В работе [Виноградов, 2015] используется следующая вероятностная модель: фантомное сходство  $(a_1, \dots, a_n) \in \{0,1\}^N$ , порождаемое из  $b$  обучающих примеров (родителей), является последовательностью испытаний Бернулли с вероятностью успеха  $a_j = 1$  равной  $(p_j)^b$ .

С помощью этой вероятностной модели теоретически было доказано: вероятность того, что возникнет фантомное сходство, которое не устранится ни одним из  $m$  контр-примеров, равна  $\sum_{j=0}^m \binom{m}{j} \cdot (-1)^j \cdot (1 - p^2 + p^{2+j})^n$ .

В следующей таблице представлены результаты работы системы по поиску фантомных гипотез, где в последних столбцах указано количество подозрительных гипотез при порогах на конкурентов равных 2, 5 и 10 соответственно.



	Обучающая выборка	Положительное свойство	Число контр-примеров	Общее число ДСМ-гипотез	Результаты (пороги на конкурентов)		
					=2	=5	=10
1	MUSHROOMS.train	быть съедобным	1944	97564	<b>2935</b>	<b>2935</b>	<b>2930</b>
2	MUSHROOMS.train	быть ядовитым	2088	59634	<b>2049</b>	<b>2047</b>	<b>2047</b>
3	MUSHROOMS.rest	быть съедобным	1972	95367	<b>2975</b>	<b>2975</b>	<b>2974</b>
4	MUSHROOMS.rest	быть ядовитым	2120	60953	<b>1994</b>	<b>1994</b>	<b>1992</b>
5	agaricus-lepiota.data	быть съедобным	3916	116579	<b>3349</b>	<b>3349</b>	<b>3349</b>
6	agaricus-lepiota.data	быть ядовитым	4208	71546	<b>1524</b>	<b>1522</b>	<b>1522</b>
7	agaricus-lepiota.data(120)	быть съедобным	17	3088	<b>378</b>	<b>378</b>	<b>366</b>
8	agaricus-lepiota.data(120)	быть ядовитым	103	110	<b>22</b>	<b>22</b>	<b>2</b>

В работе [Виноградов, 2017] была выведена асимптотическая формула при  $n \rightarrow \infty$  для вычисления вероятности появления фантомного сходства, избегающего  $m = b \cdot \sqrt{n}$  случайных контр-примеров, если наличие атрибута как в обучающем примере, так и в контр-примере возникает как испытание Бернулли с вероятностью успеха  $p = \sqrt{\frac{a}{n}}$ .

При числе атрибутов  $n \rightarrow \infty$  и вероятности появления атрибута у контр-примеров и обучающих примеров равной  $p = \sqrt{\frac{a}{n}}$  вероятность возникновения фантомного сходства, не устраненного никаким из  $m = b \cdot \sqrt{n}$  контр-примеров, будет стремиться к  $1 - e^{-a} - a \cdot e^{-a} \cdot [1 - e^{-b\sqrt{a}}]$ .

Можно использовать эту формулу как оценку среднего числа «фантомных гипотез», так как по обучающей выборке и множеству контр-примеров можно оценить вероятность появления каждого признака (как случайную величину Бернулли).

После теоретического обоснования было решено вычислить асимптотическую формулу по средней вероятности появления признака (сокращенно по средней) и по наименьшей ненулевой вероятности появления признака (сокращенно по минимуму), а затем сравнить расчетные вероятности с экспериментальной. С результатами вычислений можно ознакомиться в таблице:

	Обучающая выборка	Положительное свойство	Число контр-примеров	Общее число ДСМ-гипотез	Результаты		
					Экспериментальная вер-ть	Расчетная вер-ть по средней	Расчетная вер-ть по минимуму
1	MUSH-ROOMS.train	быть съедобным	1944	97564	<b>0.0300828</b>	<b>0.985453</b>	<b>1.39456e-05</b>
2	MUSH-ROOMS.train	быть ядовитым	2088	59634	<b>0.0343596</b>	<b>0.985453</b>	<b>1.25289e-05</b>
3	MUSH-ROOMS.rest	быть съедобным	1972	95367	<b>0.0311953</b>	<b>0.985326</b>	<b>3.94656e-06</b>
4	MUSH-ROOMS.rest	быть ядовитым	2120	60953	<b>0.0327137</b>	<b>0.985326</b>	<b>3.80637e-06</b>
5	agaricus-lepiota.data	быть съедобным	3916	116579	<b>0.0335909</b>	<b>0.985389</b>	<b>3.77244e-06</b>
6	agaricus-lepiota.data	быть ядовитым	4208	71546	<b>0.0213009</b>	<b>0.985389</b>	<b>3.2673e-06</b>

Взглянув на выделенные столбцы и внимательно их изучив, мы видим, что значения расчетной вероятности по средней близки к 1, в то время как значения экспериментальной вероятности – к 0.

Анализ распределения признаков в обучающей выборке (после их кодирования) показывает, что частоты их появления крайне неоднородны. По причине использования асимптотической формулы Д.В. Виноградова, оперирующей с постоянной вероятностью появления признака (что, очевидно, не так), результаты выглядят так обескураживающе.

В связи с этим было принято решение использовать асимптотическую оценку с вероятностью появления признака, оцениваемой по минимуму частоты. Результаты приведены в последнем столбце, которые тоже, к сожалению, являются неудовлетворительными, т.к. полученные значения расчетной вероятности по минимуму значительно меньше значений экспериментальной вероятности.

Получается, что расчет асимптотической формулы по средней вероятности появления признака и по наименьшей ненулевой вероятности появления признака практически не имеет смысла, ибо значения слишком различаются.

### **3. Экспериментальное обнаружение переобучения в ВКФ-методе**

Для преодоления трудностей, возникающих при реализации вычислений для интеллектуального анализа данных в ДСМ-методе, Д.В. Виноградовым был разработан (см., например [Vinogradov, 2018]) вероятностно-комбинаторный формальный метод, основанный на бинарной операции сходства.

ВКФ-метод использует эффективное представление объектов битовыми строками. Собирая их все вместе, мы получаем прямоугольную таблицу, заполненную нулями и единицами и называемую обучающей выборкой. Её можно воспринимать как бинарное отношение между элементами множества  $O$ , которые мы называем именами объектов, и элементами множества  $F$ , которые мы называем признаками. Если в строке, соответствующей объекту  $o \in O$ , и столбце, соответствующем фрагменту  $f \in F$ , стоит единица, то мы говорим, что объект  $o$  обладает признаком  $f$ , и обозначаем это через  $oIf$ . В противном случае, говорим, что объект  $o$  не имеет признака  $f$ .

В связи с таким представлением, мы имеем возможность реализовать операцию сходства между объектами как побитовое умножение соответствующих битовых строк, а операция сходства между признаками – это побитовое умножение соответствующих битовых столбцов.

ВКФ-метод, как и ДСМ-метод, нацелен на поиск кандидатов в гипотезы – пар  $\langle A, B \rangle$ , где  $B' = A$  и  $A' = B$ , т.е. в обучающей выборке нам необходимо искать максимальные прямоугольники, заполненные 1.

Ключевая процедура ВКФ-метода – вероятностный алгоритм нахождения сходств с помощью спаривающей цепи Маркова и операции «Замыкай-по-одному-вверх/вниз», при помощи которых выполняется переход от одного кандидата к другому.

Нами активно проводились эксперименты по обнаружению эффекта переобучения при применении ВКФ-метода. Общая обучающая выборка была разбита с помощью датчика случайных чисел на две подвыборки;

- MUSHROOMS.train содержит 4032 примера (2088 съедобных и 1944 ядовитых грибов).
- MUSHROOMS.rest содержит 4092 примера (2120 съедобных и 1972 ядовитых грибов)

Имеется возможность в ВКФ-системе менять местами как знак целевого свойства (т.е. положительными примерами могут выступать не только съедобные, но и ядовитые грибы), так и менять местами обучающую и тестовые выборки.

При проведении нескольких экспериментов (эта множественность попыток необходима из-за вероятностного характера ВКФ-метода) и запросе даже значительного числа ВКФ-гипотез (500-1000) переобучение не наблюдалось. При этом все тестовые примеры классифицировались правильно.

Эти результаты исследований показывают, что ВКФ-метод значительно в меньшей степени подвержен эффекту переобучения. Несомненно, мы не можем доказать, что в ВКФ-методе никогда этого не случится, но в наших экспериментах этого не было.

## Заключение

Проведенные эксперименты показали, что ВКФ-метод в значительно меньшей степени подвержен феномену переобучения, проявляющемуся в виде порождения «фантомных» сходств.

С другой стороны, формулы из работы [Виноградов, 2015] оказались малоприменимыми на практике из-за неодинаковости частот появления признаков.

Но самое важное, в ДСМ-методе феномен переобучения экспериментально наблюдается. И что особенно печально, имеется возможность ошибочно классифицировать ядовитые грибы как съедобные. В связи с этим вопросу выбора причины, какого знака мы ищем, следует уделять существенное внимание.

**Благодарности.** Автор выражает искреннюю благодарность своему научному руководителю Д.В. Виноградову за постановки задач, полезные обсуждения результатов и поддержку. Также хочется поблагодарить коллег из ФИЦ ИУ РАН за поддержку и конструктивные замечания.

## Список литературы

- [Виноградов, 2015] Виноградов Д.В. Вероятность порождения случайного ДСМ-сходства при наличии контр-примеров // Научная и техническая информация, Сер. 2. – 2015. – № 3.
- [Виноградов, 2017] Виноградов Д.В. Предельная вероятность порождения случайного сходства при наличии контр-примеров // Научная и техническая информация, Сер. 2. – 2017. – № 2.
- [Vinogradov, 2018] Vinogradov D.V. Machine Learning Based on Similarity Operation // Communications in Computer and Information Science. – 2018. – Vol. 934. – P. 46-59.
- [Финн и др., 2009] Финн В.К., Аншаков О.М. ДСМ-метод автоматического порождения гипотез: Логические и эпистемологические основания – М.: URSS, 2009.
- [Ganter et al., 1999] Ganter B., Wille R. Formal Concept Analysis. – В.: Springer, 1999.

## *Abstracts of Plenary Reports*

### **KNOWLEDGE AND DATA IN ARTIFICIAL INTELLIGENCE: A DUEL OR A DUO**

T.A. Gavrilova (*gavrilova@gsom.spbu.ru*)  
Graduate School of Management SPbU, Saint Petersburg, Russia

The article discusses the methodological aspects of the two main areas in the field of artificial intelligence – neural networks and semantics. The basic paradigms of the inductive and deductive approach are considered, as well as the specifics of data and knowledge as the "fuel" of intelligent technologies.

### **AUTOMATIC METHODS FOR EXTRACTING TAXONOMICAL RELATIONS FROM TEXTS**

N.V. Lukashevich (*louk\_nat@mail.ru*)  
Lomonosov Moscow State University, Research Center of Trusted Artificial Intelligence of the V.P. Ivannikov Institute of System Programming of the Russian Academy of Sciences, Moscow, Russia

This paper provides an overview of approaches to automatic extraction of taxonomic relationships between words, including classical methods based on lexico-semantic patterns, vector representations of words, and their modern development. It also describes new approaches based on language models like BERT and using new options for working with lexico-semantic templates, in which the model fills the target position in the template. The paper presents various ways to test the quality of extracting taxonomic relations and the results of these tests. According to the results currently achieved, it can be concluded that despite the progress in the development of methods for extracting taxonomic relations, the quality of extraction is insufficient for fully automatic construction or enrichment of taxonomies.

### **ARTIFICIAL INTELLIGENCE FOR CYBERSECURITY: CONFRONTATION BETWEEN ATTACK AND DEFENSE TECHNOLOGIES**

I.V. Kotenko (*ivkote1@mail.ru*)  
St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), Saint Petersburg, Russia

Cybersecurity has currently become a big social problem, and innovative solutions are needed to ensure the cybersecurity of society. Recent advances in artificial intelligence (AI) technologies have led to increased innovation and automation of information security processes. AI has emerged as a viable approach to handling huge amounts of heterogeneous data and performing fundamental cybersecurity tasks such as vulnerability management and security assessment, security monitoring, asset prioritization, distributed access control. While these technologies offer a significant increase in cybersecurity efficiency, they can also be used maliciously. Targeted cyberattacks have demonstrated the deliberate

use of AI for malicious purposes. Attackers are constantly changing and improving attack strategies and methods, with a particular focus on the use of AI techniques called AI-based cyberattacks, which can be used in combination with conventional attack techniques to cause more damage. The talk shows the state of the art in the use of AI in cybersecurity in the form of a comparison of methods and technologies of attack and defense. The analysis of the data used to solve cybersecurity problems is carried out, the existing areas of application of AI for cybersecurity are summarized and key limitations are identified. Author's research in the field of intelligent monitoring of cyber security and detection of cyber attacks are considered. This research is being supported by the grant of RSF #21-71-20078 in SPC RAS.

## **INTEGRATION OF KNOWLEDGE ENGINEERING AND MACHINE LEARNING FEATURE DESIGN FOR APPLIED ARTIFICIAL INTELLIGENCE SYSTEMS**

N.G. Yarushkina (*jng@ulstu.ru*)

Ulyanovsk State Technical University, Ulyanovsk, Russia

The report is devoted to the effective integration of knowledge engineering methods and the designing of machine learning features, evaluation of data sets, presentation of recommendation systems, predictive analytics taking into account the requirements of Friendly Artificial Intelligence. The generation of a set of training data is of fundamental importance for the successful application of machine learning in solving artificial intelligence problems. Currently, machine learning methods with human participation are being developed, which allow using the scientific potential of knowledge engineering for model design and feature construction. The report demonstrates the effectiveness of the proposed approach using examples of developed applied artificial intelligence systems.

# *Abstracts of Sectional Reports*

## *Section 1*

### *Knowledge Engineering, Cognitive Research and Intelligent Text Analysis*

#### **THE PROBLEM OF CLUSTERING OF THE ONTOLOGICAL INFRASTRUCTURE FOR THE EMPIRICAL DATA DESCRIPTION AND A PROPOSAL FOR TO SOLVE IT**

A.M. Begler (*alena.begler@gmail.com*)

Graduate School of Management SPbU, Saint Petersburg, Russia

For now, there is an established ontological infrastructure for describing empirical research data. There is an established ontological infrastructure for describing empirical research data. However, its analysis has shown that it breaks down into two relatively independent clusters, which prevents the integration of data between clusters. The paper proposes an approach that allows the description of the data regardless of which cluster they were described within. The approach is based on two principles: (1) describing information entities is done within a separate cluster, independent of the subject clusters, including not using the corresponding top-level ontologies; (2) ensuring horizontal connectivity between clusters and vertical connectivity within them. The proposed approach is implemented in a redesigned empirical ontology and its applicability for linking existing clusters is shown.

#### **ONTOLOGICAL ENGINEERING BASED ON META-ASSOCIATIVE GRAPHS**

V.V. Borisov (*vbor67@mail.ru*)

The Branch of National Research University  
“Moscow Power Engineering Institute» in Smolensk, Russia

Anton E. Misnik (*anton@misnik.by*)

Inter-state educational institution of higher education  
“Belarusian-Russian University” Belarus, Mogilev

The paper proposes a variety of metagraphs, focused on the use in ontological engineering of complex systems characterized by the complexity of the structure, multicomponent nature, the presence of functional subsystems that solve various target tasks, a large number of parameters characterizing the processes of the system. Ontological engineering aims to ensure the adoption of high-quality management decisions by increasing the level of integration of the necessary information, improving the search capabilities in databases and knowledge bases, providing the possibility of collaborative usage of the knowledge base, which is provided by a single semantic description of the applied ontology. Meta-associative graphs make it possible to provide a flexible representation of the

ontology of the subject area of a complex system, suitable for modeling system processes and ensuring the functioning of information-analytical processes in the system, minimizing, in situations where there is no need for detail, the required number of steps to obtain the necessary information, but, in need for detailed data, providing the ability to obtain all related data. The inclusion of the process component in the ontological model of the system makes it possible to increase the overall flexibility of process modeling.

### **EXTENSION OF THE METAGRAPH MODEL BASED ON THE PROTOGRAPH APPROACH**

Yu.E. Gapanyuk (*gapyu@bmstu.ru*)

Bauman Moscow State Technical University, Moscow, Russia

This paper solves the problem caused by the fact that in the previous versions of the formalization of the metagraph model, the question of how the elements of the model are combined into a single metagraph was not considered in sufficient detail. The main provisions of the annotated metagraph model are considered. The concept of a protograph is considered and its formalization is proposed on the basis of operations of non-directional fit and directional fit of protograph elements. The notion of an archigraph is formalized, the relationship between an archigraph and a coloring of a protograph is considered, and the operation of extending a protograph is considered. The connection of metagraph information element (MIE) with the protograph model is considered. A new version of the formalization of the metagraph model based on connection points is proposed.

### **SELECTION OF TOOLS FOR PREPROCESSING AND THEMATIC MODELING OF SCIENTIFIC ARTICLES FROM THE DATA LAKE**

M.M. Gayanova (*gayanova.mm@ugatu.su*)

E.Yu. Sazonova (*ekaterina\_rassadnikova@mail.ru*)

O.N. Smetanina (*smoljushka@mail.ru*)

A.K. Suleymanov (*sulaz@mail.ru*)

Ufa University of Science and Technology, Ufa, Russia

The article discusses the issues of choosing models and methods for preprocessing thematic modeling of texts. Texts are scientific articles automatically extracted from open sources and uploaded to the data lake. The results of methods and models choice are the basis for the software, which allows collecting scientific publications and storing them in a data lake, conducting text preprocessing and directly thematic modeling with the ability to visualize the results. The article includes a review of the current state of the problem, a problem statement and an approach to its solution, experimental research and the creation of a software solution based on the selected models and methods.



## **AUTOMATIC NEWS SUMMARIZATION DATASET CREATION WITH DOCUMENT CLUSTERS**

B.V. Dobrov (*dobrov\_bv@mail.ru*)

D.I. Chernyshev (*chdanorbis@yandex.ru*)

Lomonosov Moscow State University, Moscow, Russia

Collecting a summarization dataset is a costly task due to work amount and level of expertise required to write quality examples. We propose ClusterVote, a new method for automatic summarization dataset creation that accounts for domain specifics by analyzing information in related documents. The method can produce both abstractive and extractive summaries of various levels of granularity. We apply ClusterVote to collect a Telegram News\*CV summarization dataset, that contains summaries for Russian and English articles. Experiments demonstrate that the most refined version of ClusterVote produces summaries with extractive features similar to popular datasets such as CNN / Daily Mail while maintaining high factuality.

## **EXPERIENCE OF USING CONTENT PATTERNS IN THE DEVELOPMENT OF ONTOLOGIES OF SCIENTIFIC SUBJECT DOMAINS**

Yu.A. Zagorulko (*zagor@iis.nsk.su*)

G.B. Zagorulko (*gal@iis.nsk.su*)

A.P. Ershov Institute of Informatics Systems of Siberian  
Branch of RAS, Novosibirsk, Russia

Currently, ontologies are widely used to formalize and systematize knowledge and data in various subject domains. To facilitate and speed up the process of developing ontologies, ontological design patterns (ODPs) have recently been used, which are formal descriptions of solutions tested in practice for typical problems of ontological modeling. The most popular type of ODPs are content patterns, with the help of which you can get a uniform and consistent representation of all the entities of the modeled subject domain. The paper provides a brief overview ODPs, proposes a methodology for constructing ontologies of scientific subject domains (SDD), the main feature of which is the use of basic ontologies and content patterns, and also analyzes the experience of using this methodology when building ontologies of various SDDs.

## **CLASSIFICATION OF PATIENT'S EMOTIONS IN NONVERBAL COMMUNICATION BASED ON MACHINE LEARNING**

I.S. Kosacheva (*ilyastalk@bk.ru*)

O.N. Smetanina (*smoljushka@mail.ru*)

Ufa State Aviation Technical University, Ufa, Russia

This paper is devoted to solving the problem of classifying the patient's emotions in non-verbal communication based on machine learning. The article reflects the current state of the problem in the field of classification of human emotions in nonverbal communication, the statement of the problem and the results of experimental research to find the most

qualitative model that allows to classify emotions. The resulting model has an architecture consisting of two recurrent subnetworks with an attention mechanism, whose outputs are combined and fed to a full-connected classification layer. The resulting model on the validation dataset has 0.77 accuracy. The model was developed in python using the tensorflow and keras frameworks. The BlazeFace model from the MediaPipe framework was used for face image extraction.

### **AUTOMATED TEST PSYCHODIAGNOSTICS AND PROBLEMS OF MONITORING SOCIAL NETWORKS**

Y.M. Kuznetsova (*kuzjum@ya.ru*)

N.V. Chudova (*nchudova@gmail.com*)

A.A. Chuganskaya (*anfi.chuganskaya@yandex.ru*)

FRC IU RAS, Moscow, Russia

The paper presents an overview of the results of text-based automated psychodiagnostics obtained with the help of a text mining tool created at the FITZ IU RAS. A theoretical analysis of the problems of the organization of psychodiagnostic monitoring of social networks is proposed, which showed that it is required both to solve a number of theoretical issues of social psychology of large groups, and to improve the methods of intellectual text analysis, which would make it possible to set the task of organizing the study of users' emotional reactions to socially significant events.

### **FORMATION OF KNOWLEDGE BASES TAKING INTO ACCOUNT ONTOLOGICAL AGREEMENTS ON THE IACPAAS PLATFORM**

Ph.M. Moskalenko (*philipmm@iacp.dvo.ru*)

V.A. Timchenko (*vadim@dvo.ru*)

Institute of Automation and Control Processes FEB RAS,  
Vladivostok, Russia

The paper presents mechanisms for assignment and verification of ontological agreements for a digraph connected two-level model of information units. The application of the proposed approaches in the process of forming knowledge bases should allow setting such agreements and checking them without intermediaries (programmers). The implementation of the proposed solutions on the IACPaaS platform, designed for the creation and use of cloud-based artificial intelligence systems, is also described.

## **ONTOMATH<sup>EDU</sup> EDUCATIONAL MATHEMATICAL ONTOLOGY: ONTOLOGY ENGINEERING PROBLEMS**

O.A. Nevzorova (*onevzoro@gmail.com*)

M.V. Falileeva (*mmwwff@mail.ru*)

E.K. Lipachev (*elipachev@gmail.com*)

L.R. Shakirova (*liliana008@mail.ru*)

A.E. Dyupina (*anastasiya.dupina@yandex.ru*)

Kazan Federal University, Kazan, Russia

A.V. Kirillovich (*alickirillovich@gmail.com*)

Joint Supercomputer Center of the Russian Academy of Sciences, Kazan, Russia

This paper discusses the tasks of ontological engineering in the field of school education. Based on the proposed approach to ontology design in the field of school mathematics, the OntoMath<sup>Edu</sup> educational multilingual mathematical ontology has been developed for the school plane geometry course. The features of the organization of educational ontologies, reflecting the specifics of subject education in different countries, including the consideration of teaching methods and the language of instruction, are described. The specific features of the organization of the subject area make it possible to take into account, among other things, the level of training of students. To solve these problems, a new structure for organizing ontologies with educational projections, educational levels and prerequisites is proposed.

## **EVALUATION OF THE QUALITY OF CUSTOMER SERVICE BASED ON THEIR EMOTIONAL SATISFACTION USING TECHNOLOGIES OF ARTIFICIAL IMMUNE SYSTEMS**

A.I. Nuriakhmetov (*nu7530@mail.ru*)

D.R. Bogdanova (*dianochka7bog@mail.ru*)

N.I. Yusupova (*yusupova.ni@ugatu.su*)

Ufa University of Science and Technology (USATU), Ufa, Russia

The article proposes an approach to assessing customer satisfaction based on recognizing the emotional states of customers before and after the provision of services. The developed method for assessing the quality of customer service based on emotion recognition is presented. As a toolkit, the method of artificial immune systems is used to recognize emotions by facial expressions in images. In the work, it was possible to achieve a maximum accuracy of 80% for the task of recognizing Paul Ekman's basic emotions. These figures were achieved on the Cohn-Kanade+ dataset. The most popular approaches to emotion recognition in images and the key concepts of emotion typification were considered. The proposed model used a computer vision approach using facial markings. The obtained coordinates of the points of the face were converted into real features, and then, their number was reduced using principal component analysis. An automated system has been developed for assessing the quality of customer service based on their emotional satisfaction, which allows taking into account the emotional states of customers after the process of providing services.

## FROM REINFORCEMENT LEARNING TO COGNITION OF NATURE

V.G. Red'ko (*vgredko@gmail.com*)

Scientific Research Institute for System Analysis,  
Russian Academy of Sciences, Moscow, Russia

The paper analyzes the approach to the study of the processes of origin of the methods of scientific cognition in autonomous agents. The importance of internal models of autonomous agents, as well as processes of prediction by agents, is emphasized. Such internal agent models and prediction processes are used in reinforcement learning. These internal models can be considered as precursors to the models of nature used in scientific knowledge.

## ARGUMENTATION EXTRACTION BASED ON INDICATOR APPROACH

E.A. Sidorova (*lsidorova@iis.nsk.su*)

I.R. Akhmadeev (*i.r.akhmadeeva@iis.nsk.su*)

I.S. Kononenko (*irina\_k@cn.ru*)

P.M. Chagina (*p.chagina@gmail.com*)

A.P. Ershov Institute of Informatics Systems  
Russian Academy of Sciences Siberian Branch,  
Novosibirsk, Russia

The paper considers an indicator approach to the extraction of arguments found in popular science literature. The types of argumentation indicators and their relationship with a set of discursive markers are presented, and the methodology for compiling a dictionary of argumentation indicators is given. An approach to the use of argumentation indicators in deep learning methods based on the analysis of the contextual indicators is proposed. To build a training sample, for each indicator the main statement is extracted, as well as the left and right contexts, usually represented by neighbor sentences. Each set is annotated regarding the presence of argumentation in accordance with the corpus markup. To build the classifier, we used a list of 143 argumentation indicators and the annotated corpus, including 162 articles of the popular science genre, hosted on the ArgNetBank Studio web platform. In total, about 4600 learning contexts were obtained based on the corpus. The results of the above experiments of extracting argumentation based on indicators showed the better performance of the indicator-based classifier.

## GENERATING INTERPRETABLE WORD EMBEDDINGS USING SUPERCONCEPTS

M.M. Tikhomirov (*tikhomirov.mm@gmail.com*)

N.V. Lukashevich (*louk\_nat@mail.ru*)

Lomonosov Moscow State University,

Center for Trusted Artificial Intelligence ISP RAS, Moscow, Russia

This paper presents an approach to create interpretable vector representations of words, in which each component of the vector corresponds to some interpreted semantic category. To obtain such categories, a lexico-semantic resource in the form of semantic network (RuWordNet) is used, as well as a representative corpus of Russian-language texts to generate vector representations. The resulting interpretable vector representations were tested for the ability to map different models to the same vector space.

## ***Section 2***

### ***Data Mining and Explicable Artificial Intelligence***

#### **PROSPECTS FOR THE USE OF EXPLANATORY ARTIFICIAL INTELLIGENCE IN THE TASKS OF PERSONALIZED MEDICINE**

A.N. Averkin (*averkin2003@inbox.ru*)

S.A. Yarushev (*Yarushev.sa@rea.ru*)

Plekhanov Russian University of Economics, Moscow, Russia

Research in the field of explanatory artificial intelligence is actively conducted all over the world, including in Russia. The development of an explanatory apparatus is extremely necessary, since the use of artificial intelligence (AI) in critical areas, such as healthcare, imposes additional requirements on the reliability of systems of this class. It is necessary to increase the level of trust in AI by increasing the level of transparency of decisions made by AI. This paper provides an overview of the most promising areas of healthcare in which systems based on artificial intelligence can be used. Also, various neuro-fuzzy models of artificial neural networks were considered, potentially promising for the development of explanatory artificial intelligence systems based on them.

#### **EMPIRICAL RISK MINIMIZATION AS A TOOL TO CHOOSE A NUMBER OF VKF-HYPOTHESES**

D.V. Vinogradov (*krrguest@yandex.ru*)

Federal Research Center "Informatics and Management"  
of the RAS, Moscow, Russia

The paper represents a new approach to estimation of a required number of hypotheses on the target class causes. It follows to classical method of V.N. Vapnik and A.Ya. Chervonenkis to minimize errors on training sample. In the simplest case of Boolean algebra without counter-examples main result asserts very fast decreasing of prediction errors number when the volume of training sample tends to infinity.

#### **QUICK SEARCH FOR OPTIMAL SHORT CLASSIFIERS**

E.O. Dudyrev (*eo.dudyrev@hse.ru*)

S.O. Kuznetsov (*skuznetsov@hse.ru*)

National Research University Higher School of Economics, Moscow, Russia

Research in the field of Explainable Artificial Intelligence shows that the model must be small in order to be understandable to humans. Limiting the size of the model dramatically reduces the space of possible solutions. At the same time, many rule learning models still rely on greedy algorithms to create ensembles of decision trees. This paper discusses mathematical and engineering methods based on the ideas of Formal Concept Analysis for the efficient search for optimal short binary classifiers.

### **THREE QUESTIONS (FOR UNDERSTANDING), ADDRESSED TO "PARTY COMRADES**

M.I. Zabezhailo (*m.zabezhailo@yandex.ru*)

FRC "Computer Science and Control" RAS, Moscow, Russia

Some ideas regarding the "quality" of the results of intelligent data analysis (IDA) are discussed. The "quality" of the training data set, the language of knowledge presentation used and the actual "tools" of the IAD are considered. A number of questions have been formed, acceptable answers to which will increase the "quality" of the IDA.

### **ON THE ROLE OF ARGUMENTATION IN OPINIONS ANALYSIS**

M.A. Mikheyenkova (*m.mikheyenkova@yandex.ru*)

V.K. Finn (*v.k.finn@yandex.ru*)

FRC "Computer Science and Control" RAS, Moscow, Russia

RSUH, Moscow, Russia

The paper considers variants of argumentative semantics for many-valued logics of non-sense and uncertainty types. The proposed formal means are used to form opinions based on argumentation and analysis of their accordance with objective knowledge in a sociological poll situation. The dependence of the value of logical consistency of opinion in relation to such knowledge on argumentative semantics is shown, which indicates the need to choose tools adequate to the nature of the problem.

### **HYBRIDIZATION OF ONTOLOGIES AND NEURAL NETWORKS FOR THE DETECTION OF TIME SERIES ANOMALIES**

V.S. Moshkin (*postforvadim@ya.ru*)

D.S. Kurilo (*dimokkurilo@gmail.com*)

I.A. Andreev (*ares-ilya@yandex.ru*)

Ulyanovsk State Technical University, Ulyanovsk, Russia

The paper describes the results of the development of an algorithm for detecting anomalies in time series, taking into account the specifics of the subject area. The algorithm involves finding a time series forecast using LSTM networks, detecting anomalies based on the obtained forecast, filtering the found anomalies in accordance with possible deviations of the time series values from the trend reflected in the ontology, and logically deriving search results using a set of SWRL rules. The effectiveness of the proposed approach has been confirmed by a number of experiments conducted on the benchmark of data on the operation of oil rigs.

## **ON SYSTEMS OF SLICES IN THE STRUCTURE OF SUPERINTUITIONISTIC PREDICATE (OR PROPOSITIONAL) LOGICS. I**

D.P. Skvortsov (*skvortsovd@yandex.ru*)  
FRC “Computer Science and Control” RAS, Moscow, Russia

The structure (i.e., the lattice) of superintuitionistic predicate logics is considered. An approach to classifying logics from the structure is discussed; this means that suitable partitions (via corresponding equivalence relations) are introduced. A few examples of such partitions are demonstrated (or mentioned).

## **INTELLIGENT RISK ANALYSIS ON THE EXAMPLE OF THE COVID-19 EPIDEMIC**

A.V. Sokolov (*alexander.v.sokolov@gmail.com*)  
Institute for Information Transmission Problems of the RAS  
(Kharkevich Institute), Moscow, Russia

G.V. Royzenzon (*rgv@isa.ru*)  
Institute of System Analysis of Federal Research Center  
"Informatics and Management" of the RAS, Moscow, Russia

N.P. Komendantova (*komendan@iiasa.ac.at*)  
L. Ekenberg (*lovek@dsv.su.se*)  
International Institute of Applied Systems Analysis, Laxenburg, Austria

The paper proposes a scheme for constructing a unified multidimensional classification of risk analysis methods. Within the framework of the model approach of risk analysis, a technology of balanced identification is proposed. This technology served as the basis for creating a system for monitoring and predicting the state of hazardous phenomena and objects. A practical example of using the proposed approach to predict the development of the Covid-19 epidemic in Moscow (from March 2020 to September 2022) is considered. It is shown how the discrepancy between the forecast and reality leads (after a critical analysis) to a modification of the model or a revision of the accepted scenario of external influence. The prospects for the development of new methods of intellectual risk analysis are critically analyzed.

## **SUDDEN CARDIAC DEATH RISK ASSESSMENT USING MACHINE LEARNING**

N.I. Yusupova (*yusupova.ni@ugatu.su*)  
M.P. Богданов (*bogdanov\_marat@mail.ru*)  
Ufa State Aviation Technical University, Ufa, Russia

The paper is about researching a predictor of sudden cardiac death – atrial fibrillation. Public databases of digitized electrocardiograms were used. Two approaches were used: machine learning and digital biomarkers. It was found that using the Category boost



method, it is possible to detect atrial fibrillation against the background of other forms of arrhythmia and normal sinus rhythm with an efficiency of 99%. Orthogonal wavelets obtained from two-second samples of the ECG Fragment Database for the Exploration of Dangerous Arrhythmia were used as digital biomarkers. Wavelets have been shown to be highly selective for atrial fibrillation.

## **THE STUDY OF OVERFITTING IN ALGEBRAIC MACHINE LEARNING**

L.A. Iakimova (*jll0dmila@yandex.ru*)

Federal Research Center "Computer Science and Control"  
of the Russian Academy of Sciences, Moscow, Russia

This article describes the results of an experimental research of solvers (JSM and VKF systems) based on a binary similarity operation. They concern the classical problem of overfitting (especially its stability with respect to increasing the minimum number of generating examples) and the phenomenon of phantom similarities appearance, which leads to such overfitting.

## АВТОРСКИЙ УКАЗАТЕЛЬ

Аверкин А.Н.	248	Михеенкова М.А.	290
Андреев И.А.	301	Москаленко Ф.М.	157
Ахмадеева И.Р.	219	Мошкин В.С.	301
Беглер А.М.	38	Невзорова О.А.	177
Богданов М.Р.	335	Нуриахметов А.И.	189
Богданова Д.Р.	189	Редько В.Г.	208
Борисов В.В.	49	Ройзензон Г.В.	322
Виноградов Д.В.	257	Сазонова Е.Ю.	74
Гаврилова Т.А.	8	Сидорова Е.А.	219
Гапанюк Ю.Е.	62	Скворцов Д.П.	311
Гаянова М.М.	74	Сметанина О.Н.	74, 133
Добров Б.В.	96	Соколов А.В.	322
Дудырев Е.О.	266	Сулейманов А.К.	74
Дюпина А.Э.	177	Тимченко В.А.	157
Забейайло М.И.	280	Тихомиров М.М.	234
Загорулько Г.Б.	120	Фалилеева М.В.	177
Загорулько Ю.А.	120	Финн В.К.	290
Кириллович А.В.	177	Чагина П.М.	219
Комендантова Н.П.	322	Чернышев Д.И.	96
Кононенко И.С.	219	Чуганская А.А.	146
Косачев И.С.	133	Чудова Н.В.	146
Котенко И.В.	36	Шакирова Л.Р.	177
Кузнецов С.О.	266	Экенберг Л.	322
Кузнецова Ю.М.	146	Юсупова Н.И.	189, 335
Курило Д.С.	301	Якимова Л.А.	348
Липачёв Е.К.	177	Ярушев С.А.	248
Лукашевич Н.В.	20, 234	Ярушкина Н.Г.	37
Мисник А.Е.	49		

## СОДЕРЖАНИЕ

### ПЛЕНАРНЫЕ ДОКЛАДЫ

<b>Гаврилова Т.А.</b> ЗНАНИЯ И ДАННЫЕ В ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ: ДУЭЛЬ ИЛИ ДУЭТ .....	8
<b>Лукашевич Н.В.</b> АВТОМАТИЧЕСКИЕ МЕТОДЫ ИЗВЛЕЧЕНИЯ ТАКСОНОМИЧЕСКИХ ОТНОШЕНИЙ ИЗ ТЕКСТОВ.....	20
<b>Котенко И.В.</b> ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ДЛЯ КИБЕРБЕЗОПАСНОСТИ: ПРОТИВОСТОЯНИЕ ТЕХНОЛОГИЙ НАПАДЕНИЯ И ЗАЩИТЫ .....	36
<b>Ярушкина Н.Г.</b> ИНТЕГРАЦИЯ ИНЖЕНЕРИИ ЗНАНИЙ И КОНСТРУИРОВАНИЯ ПРИЗНАКОВ МАШИННОГО ОБУЧЕНИЯ ПРИ ПРОЕКТИРОВАНИИ ПРИКЛАДНЫХ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА .....	37
<b>СЕКЦИЯ 1. ИНЖЕНЕРИЯ ЗНАНИЙ, КОГНИТИВНЫЕ ИССЛЕДОВАНИЯ И ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ТЕКСТОВ</b>	
<b>Беглер А.М.</b> ПРОБЛЕМА КЛАСТЕРИЗОВАННОСТИ ОНТОЛОГИЧЕСКОЙ ИНФРАСТРУКТУРЫ ДЛЯ ОПИСАНИЯ ЭМПИРИЧЕСКИХ ДАННЫХ И ПРЕДЛОЖЕНИЕ ПО ЕЁ РЕШЕНИЮ .....	38
<b>Борисов В.В., Мисник А.Е.</b> ОНТОЛОГИЧЕСКИЙ ИНЖИНИРИНГ В СЛОЖНЫХ СИСТЕМАХ НА ОСНОВЕ МЕТА-АССОЦИАТИВНЫХ ГРАФОВ.....	49
<b>Гапанюк Ю.Е.</b> РАСШИРЕНИЕ МЕТАГРАФОВОЙ МОДЕЛИ НА ОСНОВЕ ПРОТОГРАФОВОГО ПОДХОДА .....	62

<b>Гаянова М.М., Сазонова Е.Ю., Сметанина О.Н., Сулейманов А.К.</b>	
ВЫБОР ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ ПРЕДОБРАБОТКИ И ТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ НАУЧНЫХ СТАТЕЙ ИЗ ОЗЕРА ДАННЫХ .....	74
<b>Добров Б.В., Чернышев Д.И.</b>	
МЕТОД АВТОМАТИЧЕСКОГО ПОСТРОЕНИЯ ОБУЧАЮЩИХ КОЛЛЕКЦИЙ ДЛЯ ЗАДАЧИ АБСТРАКТИВНОГО АННОТИРОВАНИЯ НОВОСТНЫХ СТАТЕЙ.....	96
<b>Загорулько Ю.А., Загорулько Г.Б.</b>	
ОПЫТ ИСПОЛЬЗОВАНИЯ ПАТТЕРНОВ СОДЕРЖАНИЯ ПРИ РАЗРАБОТКЕ ОНТОЛОГИЙ НАУЧНЫХ ПРЕДМЕТНЫХ ОБЛАСТЕЙ .....	120
<b>Косачев И.С., Сметанина О.Н.</b>	
КЛАССИФИКАЦИЯ ЭМОЦИЙ ПАЦИЕНТА ПРИ НЕВЕРБАЛЬНОЙ КОММУНИКАЦИИ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ.....	133
<b>Кузнецова Ю.М., Чудова Н.В., Чуганская А.А.</b>	
АВТОМАТИЗИРОВАННАЯ ТЕКСТОВАЯ ПСИХОДИАГНОСТИКА И ПРОБЛЕМЫ МОНИТОРИНГА СОЦИАЛЬНЫХ СЕТЕЙ .....	146
<b>Москаленко Ф.М., Тимченко В.А.</b>	
ФОРМИРОВАНИЕ БАЗ ЗНАНИЙ С УЧЁТОМ ОНТОЛОГИЧЕСКИХ СОГЛАШЕНИЙ НА ПЛАТФОРМЕ IASRAAS .....	157
<b>Невзорова О.А., Фалилеева М.В., Липачёв Е.К., Шакирова Л.Р., Дюпина А.Э., Кириллович А.В.</b>	
ОБРАЗОВАТЕЛЬНАЯ ОНТОЛОГИЯ ОНТОМАТН <sup>EDU</sup> : ПРОБЛЕМЫ ОНТОЛОГИЧЕСКОГО ИНЖИНИРИНГА .....	177
<b>Нуриахметов А.И., Богданова Д.Р., Юсупова Н.И.</b>	
ОЦЕНКА КАЧЕСТВА ОБСЛУЖИВАНИЯ КЛИЕНТОВ НА ОСНОВЕ ИХ ЭМОЦИОНАЛЬНОЙ УДОВЛЕТВОРЕННОСТИ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ....	189

<b>Редько В.Г.</b> ОТ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ К ПОЗНАНИЮ ПРИРОДЫ .....	208
<b>Сидорова Е.А., Ахмадеева И.Р., Кононенко И.С., Чагина П.М.</b> ИЗВЛЕЧЕНИЕ АРГУМЕНТАЦИИ НА ОСНОВЕ ИНДИКАТОРНОГО ПОДХОДА .....	219
<b>Тихомиров М.М., Лукашевич Н.В.</b> ПОРОЖДЕНИЕ ИНТЕРПРЕТИРУЕМЫХ ВЕКТОРНЫХ ПРЕДСТАВЛЕНИЙ СЛОВ НА ОСНОВЕ СУПЕРПОНЯТИЙ.....	234
<b>СЕКЦИЯ 2. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И ОБЪЯСНИМЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ</b>	
<b>Аверкин А.Н., Ярушев С.А.</b> ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ ОБЪЯСНИТЕЛЬНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЗАДАЧАХ ПЕРСОНАЛИЗИРОВАННОЙ МЕДИЦИНЫ .....	248
<b>Виноградов Д.В.</b> МИНИМИЗАЦИЯ ЭМПИРИЧЕСКОГО РИСКА КАК СРЕДСТВО ВЫБОРА ЧИСЛА ВКФ-ГИПОТЕЗ .....	257
<b>Дудырев Е.О., Кузнецов С.О.</b> БЫСТРЫЙ ПОИСК ОПТИМАЛЬНЫХ КОРОТКИХ КЛАССИФИКАТОРОВ .....	266
<b>Забейло М.И.</b> ТРИ ВОПРОСА (НА ПОНИМАНИЕ), АДРЕСОВАННЫЕ «ТОВАРИЩАМ ПО ПАРТИИ» .....	280
<b>Михеенкова М.А., Финн В.К.</b> О РОЛИ АРГУМЕНТАЦИИ В АНАЛИЗЕ МНЕНИЙ .....	290
<b>Мошкин В.С., Курило Д.С., Андреев И.А.</b> ГИБРИДИЗАЦИЯ ОНТОЛОГИЙ И НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ ДЕТЕКТИРОВАНИЯ АНОМАЛИЙ ВРЕМЕННЫХ РЯДОВ .....	301
<b>Скворцов Д.П.</b> О СИСТЕМАХ СЛОЁВ В СТРУКТУРЕ СУПЕРИНТУИЦИОНИСТСКИХ ПРЕДИКАТНЫХ (ИЛИ ПРОПОЗИЦИОНАЛЬНЫХ) ЛОГИК. I .....	311

<b>Соколов А.В., Ройзензон Г.В., Комендантова Н.П., Экенберг Л.</b> ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ РИСКОВ НА ПРИМЕРЕ ЭПИДЕМИИ COVID-19.....	322
<b>Юсупова Н.И., Богданов М.Р.</b> ОЦЕНКА РИСКА ВНЕЗАПНОЙ СЕРДЕЧНОЙ СМЕРТИ С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ .....	335
<b>Якимова Л.А.</b> ИССЛЕДОВАНИЕ ПЕРЕОБУЧЕНИЯ В АЛГЕБРАИЧЕСКОМ МАШИННОМ ОБУЧЕНИИ .....	348
Abstracts of Plenary Reports .....	356
Abstracts of Sectional Reports .....	358
АВТОРСКИЙ УКАЗАТЕЛЬ .....	369

*Научное издание*

Двадцатая Национальная конференция  
по искусственному интеллекту с международным участием  
КИИ-2022

Труды конференции  
Том 1

21–23 декабря 2022

---

Подписано в печать	30.11.22.	Печать ризография	Формат 60x84 1/16
Печ. л. 23,5	Тираж 90 экз.	Изд. № 22н-156	Заказ

---

Оригинал-макет подготовлен в РИО НИУ «МЭИ».  
111250, г. Москва, ул. Красноказарменная, д. 14.  
Отпечатано в типографии НИУ «МЭИ».  
111250, г. Москва, ул. Красноказарменная, д. 13.



**ДЛЯ ЗАМЕТОК**